



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 109338

TO: Manjunath N Rao
Location: CM1/10A11&10D)1
Art Unit: 1652
December 10, 2003

Case Serial Number: 10/074527

From: P. Sheppard
Location: CM1-1E03
Phone: (703) 308-4499

sheppard@uspto.gov

Search Notes

Examiner Search notes.

THIS PAGE BLANK (USPTO)

STIC-Biotech/ChemLib

109338

From: Rao, Manjunath N.
Sent: Monday, December 01, 2003 10:07 AM
T : STIC-Biotech/ChemLib
Subject: Sequence search request for 10/074,527

From: Manjunath N. Rao
Art Unit 1652, Room 10A11
Mail Box in Room 10D 01
Phone: 306-5681

CRFF

Date: 12-1-03

Please search the following as soon as possible for application with serial number

10/074,527

1. SEQ ID NO: 1 and 3 against all commercial nucleic acid databases, issued patents/published applications database and pending application database. Please provide a print of all results
2. SEQ ID NO: 2 against all commercial amino acid databases, issued patents/published applications database and pending application database. Please provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.
Biotechnology Patent Examiner
Art Unit 1652, Room 10A11
Mail Box in 10D01
Crystal Mall 1, USPTO.

Point of Contact
P. Sheppard
telephone number: (703) 308-4499

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 12/1/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

US-10-198-846-9749

Query Match 100.0%; Score 1746; DB 14; Length 2904;
Best Local Similarity 100.0%; Pred. No. 0;
Matches: 1746; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGTGGGGCGGACGGCGGGCGGCGCTGCCCGGGAACTCGCGCGCGCGCGCGCGGAGCGG	60
DB	2826	ATGTGGGGCGGACGGCGGGCGGCGCTGCCCGGGAACTCGCGCGCGCGCGCGGAGCGG	2767
QY	61	CTGTTGGTCTCTCGCGCTACTTGGCGTTGGCGGGCTGGGCTCGGTCTGGCGGCGCAG	120
DB	2766	CTGTTGGTCTCTCGCGCTACTTGGCGTTGGCGGGCTGGGCTCGGTCTGGCGGCGCAG	2707
QY	121	CGTGGGGCGGGCGGGGCTGCCGAGCGGGACCCCCCGGACCCCCCGCGCCCGGGCGG	180
DB	2706	CGTGGGGCGGGCGGGGCTGCCGAGCGGGACCCCCCGGACCCCCCGCGCCCGGGCGG	2647
QY	181	CGGAGCGGTCATCGCGGGCGCGGCTGCCGGGAAACCGCTGGCGCGGGGCGGAG	240
DB	2646	CGGAGCGGTCATCGCGGGCGCGGCTGCCGGGAAACCGCTGGCGCGGGGCGGAG	2587
QY	241	GCCTGCGGCTGCAGCTGCAGGCGGAGGCTCGGCTGCAGGAGGAGCGTGCGGCTG	300
DB	2586	GCCTGCGGCTGCAGCTGCAGGCGGAGGCTCGGCTGCAGGAGGAGCGTGCGGCTG	2527
QY	301	CACAGATTAACTTACCTCAGGACCGCATCTCTACTGACCGCGCGCTGCCCGGAGCG	360
DB	2526	CACAGATTAACTTACCTCAGGACCGCATCTCTACTGACCGCGCGCTGCCCGGAGCG	2467
QY	361	TGGAAACCGCTGCAAGAGAGAAATATGATTATGATAATTTGCCAGGACATCTGTT	420
DB	2466	TGGAAACCGCTGCAAGAGAGAAATATGATTATGATAATTTGCCAGGACATCTGTT	2407
QY	421	ATCATAGCATTTTATATGAAGCTGGTCAACTCTCTCTCGGACAGTTTACAGTCTCTT	480
DB	2406	ATCATAGCATTTTATATGAAGCTGGTCAACTCTCTCTCGGACAGTTTACAGTCTCTT	2347
QY	481	GAGACATCCCGGATATCTGCTAGAGAAAGTATCTTGTAGATGATGATGATGATGATA	540
DB	2346	GAGACATCCCGGATATCTGCTAGAGAAAGTATCTTGTAGATGATGATGATGATGATA	2287
QY	541	GAGCACCTGAAGAGCGCTTGGCAATGAGCTTTTCGGGACTGCCCAAGTGGCGCTGATC	600
DB	2286	GAGCACCTGAAGAGCGCTTGGCAATGAGCTTTTCGGGACTGCCCAAGTGGCGCTGATC	2227
QY	601	CGCGCCAAACAGAGAGAGGCGCTGGTGGAGCGCGGCTGCTGGGGCGTCTGGCGGAGG	660
DB	2226	CGCGCCAAACAGAGAGAGGCGCTGGTGGAGCGCGGCTGCTGGGGCGTCTGGCGGAGG	2167
QY	661	GGCGATGTTCTGACCTTCTGGACTGTCTGAGTGCCACGAGGGTGGCTGGAGCGG	720
DB	2166	GGCGATGTTCTGACCTTCTGGACTGTCTGAGTGCCACGAGGGTGGCTGGAGCGG	2107
QY	721	CTGCTGCAGAGGATCCATGAAGAGAGTGGGAGTGGTGTGCCCGGTGATTGATGATC	780
DB	2106	CTGCTGCAGAGGATCCATGAAGAGAGTGGGAGTGGTGTGCCCGGTGATTGATGATC	2047
QY	781	GACTGGAAACACCTTTCGAATACCTGGGAACTCCGGGAGCGCCAGATCGCGGTTTCGAC	840
DB	2046	GACTGGAAACACCTTTCGAATACCTGGGAACTCCGGGAGCGCCAGATCGCGGTTTCGAC	1987
QY	841	TGGAGGCTGGTTTACCTGGGACACAGTTTCTGAGGGGAGAGATACGATCAATCC	900
DB	1986	TGGAGGCTGGTTTACCTGGGACACAGTTTCTGAGGGGAGAGATACGATCAATCC	1927
QY	901	CCCGTCGATGTCATCAGGTCCTCAACAATGGCTGGTGGGCTGTTTGTGTGAGTAAGAAA	960
DB	1926	CCCGTCGATGTCATCAGGTCCTCAACAATGGCTGGTGGGCTGTTTGTGTGAGTAAGAAA	1867
QY	961	TATTTTGAATATCTGGGGCTTTATGATACAGAAATGGAAGTTTGGGGAGGAGAAACCTC	1020
DB	1866	TATTTTGAATATCTGGGGCTTTATGATACAGAAATGGAAGTTTGGGGAGGAGAAACCTC	1807

QY	1021	GAATTTTCTTTAGGATCTGCAGTGTGGTGTCTGGAAACACACCCATGTTCCCAT	1080
DB	1806	GAATTTTCTTTAGGATCTGCAGTGTGGTGTCTGGAAACACACCCATGTTCCCAT	1747
QY	1081	GTGGCCATGTTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAAGGCTCTGGGCAACAGT	1146
DB	1746	GTGGCCATGTTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAAGGCTCTGGGCAACAGT	1687
QY	1141	GTTGCTGAGCTGAAGTATGATGATGAATTTTAAAGAGCTCTACTACCATCGCAACCCC	1200
DB	1686	GTTGCTGAGCTGAAGTATGATGATGAATTTTAAAGAGCTCTACTACCATCGCAACCCC	1627
QY	1201	CGTCCCGCTTGGAACTTTTGGGAGTGCACAGAGAGGAGCAAGCTCCGGGCAACAGCTC	1260
DB	1626	CGTCCCGCTTGGAACTTTTGGGAGTGCACAGAGAGGAGCAAGCTCCGGGCAACAGCTC	1567
QY	1261	CAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGATGTGCTGAG	1320
DB	1566	CAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGATGTGCTGAG	1507
QY	1321	GACAGGCTGGCTTCTTGGGAGTGCACAGAAAGGACTTAACAGACTACTGCTTTGAC	1380
DB	1506	GACAGGCTGGCTTCTTGGGAGTGCACAGAAAGGACTTAACAGACTACTGCTTTGAC	1447
QY	1381	TATTAACCTCCGATGAAACACAGATTGTGGGACACACAGGTCTTCTGTACTCTGTCT	1440
DB	1446	TATTAACCTCCGATGAAACACAGATTGTGGGACACACAGGTCTTCTGTACTCTGTCT	1387
QY	1441	GGGATGGGCCAGAACTAGTTTTTCAGTACACGTCCCAAGAAAGAAATACGCTATAACACC	1500
DB	1386	GGGATGGGCCAGAACTAGTTTTTCAGTACACGTCCCAAGAAAGAAATACGCTATAACACC	1327
QY	1501	CACAGCTGAGGCTGCAATGCTGTGGAACGGAATGGAATACCTTATCATGATCTC	1560
DB	1326	CACAGCTGAGGCTGCAATGCTGTGGAACGGAATGGAATACCTTATCATGATCTC	1267
QY	1561	TGCGAAGAAACTGCCCCAGAGAACTCAGAAATGTTTCTTTCAGAGGAGTGGATCTTTATTT	1620
DB	1266	TGCGAAGAAACTGCCCCAGAGAACTCAGAAATGTTTCTTTCAGAGGAGTGGATCTTTATTT	1207
QY	1621	CACGAACAGTCCAGAAATGTTGCTGAGGAGGAGGAGTGCAGTGCAGTTTCGTT	1680
DB	1206	CACGAACAGTCCAGAAATGTTGCTGAGGAGGAGGAGTGCAGTGCAGTTTCGTT	1147
QY	1681	CCACTCTTACGAGCTGCACCAACTCGGATCATCAAAATGTTCTTCAAGAGCGCATG	1740
DB	1146	CCACTCTTACGAGCTGCACCAACTCGGATCATCAAAATGTTCTTCAAGAGCGCATG	1087
QY	1741	TTATGA 1746	
DB	1086	TTATGA 1081	

RESULT 4
US-10-292-896-57
; Sequence 57, Application US/10292896
; Publication No. US20030186850A1
; GENERAL INFORMATION:
; APPLICANT: HASSAN, Helle
; APPLICANT: REIS, Celso A.
; APPLICANT: BENNETT, Eric P.
; APPLICANT: CLAUSEN, Henrik
; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GAI
; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS
; FILE REFERENCE: 4305/1H154-US3
; CURRENT APPLICATION NUMBER: US/10/292,896
; PRIOR FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/425,204
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/DK01/00328
; PRIOR FILING DATE: 2001-05-10

THIS PAGE BLANK (USPTO)

```

; PRIOR APPLICATION NUMBER: US 60/203,331
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
; LENGTH: 1746
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-292-896-57

```

[illegible]

```

RESULT 5
US-09-925-299-219
; Sequence 219, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883

```

THIS PAGE BLANK (USPTO)

Db 121 WNPCKEKKYDYNLPTSVIIAFYNEAMSTLLRTVTVSVLETSPIILLSEVILVDDYSR 180
Qy 181 EHLKERLANELSGLPKVLIRANKREGVLRARLLGASAARGDVLTFDCHCECHGWLEP 240
Db 181 EHLKERLANELSGLPKVLIRANKREGVLRARLLGASAARGDVLTFDCHCECHGWLEP 240
Qy 241 LLQRIHEESAVVCPVIDWNTFEYLGNSGSPQIGGFDWRLVFTWHTVPERIRMQS 300
Db 241 LLQRIHEESAVVCPVIDWNTFEYLGNSGSPQIGGFDWRLVFTWHTVPERIRMQS 300
Qy 301 PVDVIRPTWAGGLFAVSKYFEYLGSDYDTGMEVWGGNLEFSPRIWQCGVLETHPCSH 360
Db 301 PVDVIRPTWAGGLFAVSKYFEYLGSDYDTGMEVWGGNLEFSPRIWQCGVLETHPCSH 360
Qy 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHHRNPRARLEPFGDVTERRKQRLDKL 420
Db 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHHRNPRARLEPFGDVTERRKQRLDKL 420
Qy 421 QCKDFKWFLETVPYELHVPEDRPGFGLQNLQKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
Db 421 QCKDFKWFLETVPYELHVPEDRPGFGLQNLQKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
Qy 481 GMGQNOFFEYTSQKEIRYNTHQPEGCIATVEAGMDTLIMHLCETAPENQKFILOEDGSLF 540
Db 481 GMGQNOFFEYTSQKEIRYNTHQPEGCIATVEAGMDTLIMHLCETAPENQKFILOEDGSLF 540
Qy 541 HEQSKKCVQAARKESDSSFVPLLRDCTNSDHQKWFKKERM 581
Db 541 HEQSKKCVQAARKESDSSFVPLLRDCTNSDHQKWFKKERM 581

RESULT 2

US-10-292-896-58
; Sequence 58, Application US/10292896
; Publication No. US20030186850A1

GENERAL INFORMATION:

; APPLICANT: HASSAN, Helle
; APPLICANT: REIS, Celso A.
; APPLICANT: BENNETT, Eric P.
; APPLICANT: CLAUSEN, Henrik
; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GA
; TITLE OF INVENTION: TRANSPERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS
; TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICAMENTS
; FILE REFERENCE: 4305/1H154-US3
; CURRENT APPLICATION NUMBER: US/10/292,896
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/425,204
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/DK01/00328
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 60/203,331
; PRIOR FILING DATE: 2000-05-11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 581
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-292-896-58

Query Match 98.5%; Score 3076; DB 12; Length 581;
Best Local Similarity 98.5%; Pred. No. 2.4e-278;
Matches 572; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MWGTRARRCPRELRRGREALLVALLLAGLSVLRQAGAGAAEPGPPRTPRGR 60
Db 1 MWGTRARRCPRELRRGREALLVALLLAGLSVLRQAGAGAAEPGPPRTPRGR 60
Qy 61 REPVMPRPVPAALGARGAVRLQOGEELRLQESVRLHQINIVLSRISLHRLRPER 120
Db 61 REPVMPRPVPAALGARGAVRLQOGEELRLQESVRLHQINIVLSRISLHRLRPER 120
Qy 121 WNPCKEKKYDYNLPTSVIIAFYNEAMSTLLRTVTVSVLETSPIILLSEVILVDDYSR 180

Db 121 WNPCKEKKYDYNLPTSVIIAFYNEAMSTLLRTVTVSVLETSPIILLSEVILVDDYSR 180
Qy 181 EHLKERLANELSGLPKVLIRANKREGVLRARLLGASAARGDVLTFDCHCECHGWLEP 240
Db 181 EHLKERLANELSGLPKVLIRANKREGVLRARLLGASAARGDVLTFDCHCECHGWLEP 240
Qy 241 LLQRIHEESAVVCPVIDWNTFEYLGNSGSPQIGGFDWRLVFTWHTVPERIRMQS 300
Db 241 LLQRIHEESAVVCPVIDWNTFEYLGNSGSPQIGGFDWRLVFTWHTVPERIRMQS 300
Qy 301 PVDVIRPTWAGGLFAVSKYFEYLGSDYDTGMEVWGGNLEFSPRIWQCGVLETHPCSH 360
Db 301 PVDVIRPTWAGGLFAVSKYFEYLGSDYDTGMEVWGGNLEFSPRIWQCGVLETHPCSH 360
Qy 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHHRNPRARLEPFGDVTERRKQRLDKL 420
Db 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHHRNPRARLEPFGDVTERRKQRLDKL 420
Qy 421 QCKDFKWFLETVPYELHVPEDRPGFGLQNLQKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
Db 421 QCKDFKWFLETVPYELHVPEDRPGFGLQNLQKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
Qy 481 GMGQNOFFEYTSQKEIRYNTHQPEGCIATVEAGMDTLIMHLCETAPENQKFILOEDGSLF 540
Db 481 GMGQNOFFEYTSQKEIRYNTHQPEGCIATVEAGMDTLIMHLCETAPENQKFILOEDGSLF 540
Qy 541 HEQSKKCVQAARKESDSSFVPLLRDCTNSDHQKWFKKERM 581
Db 541 HEQSKKCVQAARKESDSSFVPLLRDCTNSDHQKWFKKERM 581

RESULT 3

US-10-074-527-8
; Sequence 8, Application US/10074527
; Publication No. US20020142426A1

GENERAL INFORMATION:

; APPLICANT: Olandt, Peter J.
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Galvin, Katherine A.
; APPLICANT: Millennium Pharmaceuticals Inc.
; TITLE OF INVENTION: 33945, A Human Glycosyltransferase and
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: MPI2001-018P1RCPI(M)
; CURRENT APPLICATION NUMBER: US/10/074,527
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/269202
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-074-527-8

Query Match 53.6%; Score 1675; DB 14; Length 578;
Best Local Similarity 57.9%; Pred. No. 2e-147;
Matches 330; Conservative 78; Mismatches 144; Indels 18; Gaps 8;

Qy 23 VILLALLAGL-----GSVLRAQAGAGAAEPGPPRTPRGRRE-----PVMPPRPVPA 72
Db 13 LLLALLTAYITLVFVSSTLYASPGAG-GARELGRRLPDLDTRFEEDLSQPLYTKPPADS 71
Qy 73 NALGARGAVRLQOGEELRLQESVRLHQINIVLSRISLHRLPERWNPCKEKKYDY 132
Db 72 HALGEWGRASKQNEGELKQOEEELIERYAINIVLSRISLHRIEDKRYTECKAKPHY 131
Qy 133 DNLPRTSVIIAFYNEAMSTLLRTVTVSVLETSPIILLSEVILVDDYSRDLHRLKERNL 192
Db 132 RSLPTTSVIIAFYNEAMSTLLRTVTVSVLETSPIILLSEVILVDDYSRILYKAQLETVIS 191
Qy 193 GLPKVRLIRANKREGVLRARLLGASAARGDVLTFDCHCECHGWLEPFLQRIHEESAV 252

THIS PAGE BLANK (USPTO)

XX The invention relates to an isolated human drug metabolising enzyme (DME)
CC and its nucleotide. DME is useful for diagnosing, treating or preventing
CC disorders associated with aberrant expression of DME, where the disorders
CC are selected from autoimmune/inflammatory disorder such as acquired
CC immunodeficiency syndrome (AIDS), asthma, atherosclerosis, psoriasis,
CC uveitis; a cell proliferative disorder such as arteriosclerosis,
CC cirrhosis, hepatitis, and cancer; a neurological disorder such as
CC Alzheimer's disease, Huntington's disease, dementia, Parkinson's disease;
CC a developmental disorder such as renal tubular acidosis, epilepsy,
CC anaemia; an endocrine disorder such as adenoma, thrombosis and
CC infections; an eye disorder such as conjunctivitis, glaucoma, cataract;
CC metabolic disorder such as cystic fibrosis, diabetes and goitre; a
CC gastrointestinal disorder such as anorexia, peptic ulcer; and liver
CC disorders. DME is useful in a number of drug screening techniques and to
CC analyse the proteome of a tissue or cell type. The invention is useful
CC for creating knock-in humanised animals or transgenic animals to model
CC human diseases, in somatic or germline gene therapy, to generate a
CC transcript image of a tissue or cell type, for detecting differences in
CC the chromosomal location due to translocation, inversion, etc. among
CC normal, carrier or affected individuals, and as hybridisation probes for
CC mapping naturally occurring genomic sequences. The present sequence is
CC human DME-4.
XX

SQ Sequence 581 AA;

Query Match 100.0%; Score 3124; DB 23; Length 581;
Best Local Similarity 100.0%; Pred. No. 8.6e-292;
Matches 581; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNGTARRCPELRGREALVLLALLAGLSVLAQRGAGAAEPPGPRTPRGR 60
DB 1 MNGTARRCPELRGREALVLLALLAGLSVLAQRGAGAAEPPGPRTPRGR 60
QY 61 REPVMPRPVPANALGARGEAVRLQOGEELRQESVRLHQINYLSDRSLHRLRPER 120
DB 61 REPVMPRPVPANALGARGEAVRLQOGEELRQESVRLHQINYLSDRSLHRLRPER 120
QY 121 WNPLCKEKYDYNLPRTSVIIAFYNEAWSTLLRTVYSLVLETSPIILLEEVILVDYSDR 180
DB 121 WNPLCKEKYDYNLPRTSVIIAFYNEAWSTLLRTVYSLVLETSPIILLEEVILVDYSDR 180
QY 181 EHLKERLANELSGLPKVLIRANKREGLVRALLGASAAARGDVLTFDCHCEHGWLEP 240
DB 181 EHLKERLANELSGLPKVLIRANKREGLVRALLGASAAARGDVLTFDCHCEHGWLEP 240
QY 241 LLQRIHEESAVVCPVIDVNTFEYLGNSGEPOIGGDFMLVFTWHTVPERIRMQS 300
DB 241 LLQRIHEESAVVCPVIDVNTFEYLGNSGEPOIGGDFMLVFTWHTVPERIRMQS 300
QY 301 PVDVIRSPMTMAGGLFAVSKKYFEYLGSDYDTGMEVNGGENLEFSFRIWQCGGVLETHPCSH 360
DB 301 PVDVIRSPMTMAGGLFAVSKKYFEYLGSDYDTGMEVNGGENLEFSFRIWQCGGVLETHPCSH 360
QY 361 VGHVFPKQAPYGRNKAANSVRAAEVWMDFFELYHNRNPRARLPFGDVTERRKQRL 420
DB 361 VGHVFPKQAPYGRNKAANSVRAAEVWMDFFELYHNRNPRARLPFGDVTERRKQRL 420
QY 421 QCKDFKWFLETYVPELHVPEDRPGFMQLQNGKLTDCFDYNNPDENQIVGHQVILYLCH 480
DB 421 QCKDFKWFLETYVPELHVPEDRPGFMQLQNGKLTDCFDYNNPDENQIVGHQVILYLCH 480
QY 481 GMGNQOFFEYTSQKEIRYNTHOPEGCIAVEAGMDTLIMHLCBETAPENKFILQEDGSIF 540
DB 481 GMGNQOFFEYTSQKEIRYNTHOPEGCIAVEAGMDTLIMHLCBETAPENKFILQEDGSIF 540
QY 541 HGSQKCKVQAARKESSDSFVPLLRDCTNSDHDQKWFPERML 581
DB 541 HGSQKCKVQAARKESSDSFVPLLRDCTNSDHDQKWFPERML 581

RESULT 3
ABG04767

ID ABG04767 standard; Protein; 355 AA.
XX
AC ABG04767;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #4758.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX human supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN W0200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR N-PSDB; AAS68954.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
PS Claim 20; SEQ ID No 35126; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIP
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 355 AA;

Query Match 57.2%; Score 1788; DB 22; Length 355;
Best Local Similarity 98.5%; Pred. No. 2.5e-163;
Matches 335; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 65 MPRPPVPANALGARGEAVRLQOGEELRQESVRLHQINYLSDRSLHRLRPERNPL 124
DB 1 MPRPPVPANALGARGEAVRLQOGEELRQESVRLHQINYLSDRSLHRLRPERNPL 60
QY 125 CKEKKYDYNLPRTSVIIAFYNEAWSTLLRTVYSLVLETSPIILLEEVILVDYSDRHLK 184
DB 61 CKEKKYDYNLPRTSVIIAFYNEAWSTLLRTVYSLVLETSPIILLEEVILVDYSDRHLK 120
QY 185 ERLANELSGLPKVLIRANKREGLVRALLGASAAARGDVLTFDCHCEHGWLEPQLQR 244

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 12:45:48 ; Search time 6368.95 seconds
(without alignments)
11215.070 Million cell updates/sec

Title: US-10-074-527-3

Perfect score: 1746

Sequence: 1 atgtggtggggcgacggcg.....tcaagagcgcatgttatga 1746

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_man.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score or the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1746	100.0	2745	6	AX477705	AX477705 Sequence
2	1744.4	99.9	1746	9	AB078146	AB078146 Homo sapi
3	1744.4	99.9	1746	9	HS132365	HS132365 Homo sapi
4	1319	75.5	2343	9	AK024865	AK024865 Homo sapi
5	851.4	48.8	63632	2	AC132018	Continuation (5 of
6	851.4	48.8	110000	2	AC132018	Continuation (3 of
7	851.4	48.8	288118	2	AC129139	AC129139 Rattus no
8	842.6	48.3	228368	2	AC121727	AC121727 Rattus no
9	841	48.2	94616	2	AC098559	Continuation (8 of
10	841	48.2	288118	2	AC129139	AC129139 Rattus no
11	830	47.5	1928	9	BC013945	BC013945 Homo sapi
12	812.6	46.5	110000	2	AC099237	Continuation (4 of
13	812.6	46.5	166547	2	AC134060	AC134060 Rattus no
14	805.4	46.1	166547	2	AC134060	AC134060 Rattus no
15	803.8	46.0	110000	2	AC098559	Continuation (2 of
16	795.4	45.6	110000	2	AC098559	AC098559 Rattus no
17	795.4	45.6	110000	2	AC112324	AC112324 Rattus no
18	795.4	45.6	110000	2	AC112324	Continuation (3 of
19	795.4	45.6	299134	2	AC128783	AC128783 Rattus no
20	792.2	45.4	110000	2	AC105836	Continuation (2 of
21	526.2	30.1	328901	2	AC130864	AC130864 Rattus no
22	522.4	29.9	226623	2	AC097343	AC097343 Rattus no
23	520.8	29.8	1737	10	MMU73819	U73819 Mus musculu
24	520.8	29.8	234109	2	AC101864	AC101864 Mus muscu
25	510.6	29.2	110000	2	AC099237	AC099237 Rattus no
26	499.4	28.6	5369	9	BC036390	BC036390 Homo sapi
27	499.4	28.6	73481	9	AC025034	AC025034 Homo sapi
28	496.2	28.4	1737	6	AR236647	AR236647 Sequence
29	496.2	28.4	1737	9	HSY08564	Y08564 Homo sapien
30	493.6	28.3	165175	2	AC083808	AC083808 Homo sapi
31	419	24.0	554	6	AX197983	AX197983 Sequence
32	418.4	24.0	544	6	AX208574	AX208574 Sequence
33	400.8	23.0	652	6	AX197725	AX197725 Sequence
34	400.2	22.9	632	6	AX208332	AX208332 Sequence
35	400.2	22.9	632	6	AX209583	AX209583 Sequence
36	374	21.4	149930	9	AL136084	AL136084 Human DNA
37	363	20.8	483	6	AX371123	AX371123 Sequence
38	300.8	17.2	2384	3	AK112431	AK112431 Ciona int
39	285.6	16.4	1920	6	AR153422	AR153422 Sequence
40	285.6	16.4	2575	6	AX376296	AX376296 Sequence
41	285.6	16.4	2575	6	AX697278	AX697278 Sequence
42	285.6	16.4	3357	9	AY035399	AY035399 Homo sapi
43	285.6	16.4	3580	9	HSN803261	AL81925 Homo sapi
44	284	16.3	3332	6	AR155987	AR155987 Sequence
45	284	16.3	3528	9	BC014789	BC014789 Homo sapi

ALIGNMENTS

RESULT 1
AX477705
LOCUS
DEFINITION Sequence 17 from Patent WO0246426.
ACCESSION AX477705
VERSION AX477705.1
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Sanjanwala, M.M., Yao, M.G., Au-Young, J., Baughn, M.R., Arvizu, C.,
Ring, H.Z., Lee, E.A., Ding, L., Hafalia, A.J., Tang, Y.T., Yue, H.,
Tribouley, C.M., Lu, D.A., Lai, P.G., Warren, B.A., Yang, J.,

2745 bp DNA linear PAT 12-AUG-2002

THIS PAGE BLANK (USPTO)

Walia, N.K., Nguyen, D.B., Gandhi, A.R. and Ison, C.H.

Drug metabolizing enzymes
Patent: WO 0246426-A 17-13-JUN-2002;
Incyte Genomics, Inc. (US)

FEATURES

Location/Qualifiers
1..2745
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 2860635CB1"

BASE COUNT 704 a 612 c 737 g 692 t

ORIGIN

Query Match 100.0%; Score 1746; DB 6; Length 2745;
Best Local Similarity 100.0%; Pred. No. 2e-266;
Matches 1746; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGGGGGCGACGCGCGGGCGCTGCCCGCGGAACTGCGGCGCGCGCGGAGCG 60
DB 1 ATGTGGGGGCGACGCGCGGGCGCTGCCCGCGGAACTGCGGCGCGCGGAGCG 60

QY 61 CTGTTGCTGCTCTGCGGCTACTGCGGCTTGGCGGCTGGGCTCGGTGCTGCGGGCGAG 120
DB 61 CTGTTGCTGCTCTGCGGCTACTGCGGCTTGGCGGCTGGGCTCGGTGCTGCGGGCGAG 120

QY 121 CGTGGGCGCGGGCGCGCGGCTGCCGAGCCGGACCCCGCGCACCCCGCGCCCGCGGCGG 180
DB 121 CGTGGGCGCGGGCGCGGCTGCCGAGCCGGACCCCGCGCACCCCGCGCCCGGGCGG 180

QY 181 CGGAGCCGGTATGATCGCGCGCGCGCGGTGCCGGGAAACGCGCTGGCGCGCGGGCGAG 240
DB 181 CGGAGCCGGTATGATCGCGCGCGCGCGGTGCCGGGAAACGCGCTGGCGCGCGGGCGAG 240

QY 241 GCGGTGCGGCTGCGAGCTGCGAGCGCGAGGAGCTGCGGCTGCGAGGAGAGCGTGGCGGTG 300
DB 241 GCGGTGCGGCTGCGAGCTGCGAGCGCGAGGAGCTGCGGCTGCGAGGAGAGCGTGGCGGTG 300

QY 301 CACCAAGATTAACTCTACCTGAGGACCGCATCTCACTGCACCCCGCGCTGCCCGAGCGC 360
DB 301 CACCAAGATTAACTCTACCTGAGGACCGCATCTCACTGCACCCCGCGCTGCCCGAGCGC 360

QY 361 TGGAAACCCGCTGCAAGAGAGAGAAATATGATTAATGATTAATTTGCCAGGACATCTGT 420
DB 361 TGGAAACCCGCTGCAAGAGAGAGAAATATGATTAATGATTAATTTGCCAGGACATCTGT 420

QY 421 ATCATAGCATTTTATTAAGAACCTGGTCAACTCTTCCCTTCGAGAGAGTTTACAGTGTCTT 480
DB 421 ATCATAGCATTTTATTAAGAACCTGGTCAACTCTTCCCTTCGAGAGAGTTTACAGTGTCTT 480

QY 481 GAGACATCCCGGATATCTGCTAGAGAGAGTATCTTTGTAGATGACTACAGTGATAGA 540
DB 481 GAGACATCCCGGATATCTGCTAGAGAGAGTATCTTTGTAGATGACTACAGTGATAGA 540

QY 541 GAGCACTTGAAGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGGTGCGCTGATC 600
DB 541 GAGCACTTGAAGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGGTGCGCTGATC 600

QY 601 CCGGCAACAGAGAGAGGCGCTGCTGCGAGCCCGGCTGCGGCGGCTGCGGCGGAGG 660
DB 601 CCGGCAACAGAGAGAGGCGCTGCTGCGAGCCCGGCTGCGGCGGCTGCGGCGGAGG 660

QY 661 GCGGATGTTCTGACCTTCTGAGTGTCTCACTGTGAGTGCACGAAGGCTGGAGCGG 720
DB 661 GCGGATGTTCTGACCTTCTGAGTGTCTCACTGTGAGTGCACGAAGGCTGGAGCGG 720

QY 721 CTGCTGCAAGAGATCCATGAAGAGAGTCCGAGTGTGTCGCCGGTGATTTGATGATC 780
DB 721 CTGCTGCAAGAGATCCATGAAGAGAGTCCGAGTGTGTCGCCGGTGATTTGATGATC 780

QY 781 CACTGGAACACTTCCGAATCTCGGGAACCTCCGGGAGCCCGAGATCGGCGGTTTCGAC 840
DB 781 CACTGGAACACTTCCGAATCTCGGGAACCTCCGGGAGCCCGAGATCGGCGGTTTCGAC 840

QY 841 TGGAGGCTGGTGTTCACGTGGCACACAGTTCCTTGAGAGGAGAGATACGATCAATCC 900
DB 841 TGGAGGCTGGTGTTCACGTGGCACACAGTTCCTTGAGAGGAGAGATACGATCAATCC 900

QY 901 CCCGTCGATGTCATCAGGCTCTCCAAATGGCTGGTGGCTGTTGCTGTGAGTAAGAAA 960
DB 901 CCCGTCGATGTCATCAGGCTCTCCAAATGGCTGGTGGCTGTTGCTGTGAGTAAGAAA 960

QY 961 TATTTTGAATATCTGGGGTCTTATGATACAGGAATGAAAGTTTGGGAGGAGAGAAACCTC 1020
DB 961 TATTTTGAATATCTGGGGTCTTATGATACAGGAATGAAAGTTTGGGAGGAGAGAAACCTC 1020

QY 1021 GAATTTTCTTTAGGATCTGGAGTGTGGGTTCTGGAACACACCCCATGTTCCCAT 1080
DB 1021 GAATTTTCTTTAGGATCTGGAGTGTGGGTTCTGGAACACACCCCATGTTCCCAT 1080

QY 1081 GTTGGCCATGTTTTCCTCCCAAGAGCTCCCTACTCCCGCAACAGGCTCTGGCCCAACAGT 1140
DB 1081 GTTGGCCATGTTTTCCTCCCAAGAGCTCCCTACTCCCGCAACAGGCTCTGGCCCAACAGT 1140

QY 1141 GTTCGTGACGCTGAAGTATGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCC 1200
DB 1141 GTTCGTGACGCTGAAGTATGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCC 1200

QY 1201 CGTCCCGCTTGGAACTTTTGGGGATGTCAGAGAGGAGAGCTCCGGGACAAAGCTC 1260
DB 1201 CGTCCCGCTTGGAACTTTTGGGGATGTCAGAGAGGAGAGCTCCGGGACAAAGCTC 1260

QY 1261 CAGTGTAAAGACTTCAAGTGTCTTGGGAGCTGTATCCAGAACGTCAGTGTGCTTGAC 1320
DB 1261 CAGTGTAAAGACTTCAAGTGTCTTGGGAGCTGTATCCAGAACGTCAGTGTGCTTGAC 1320

QY 1321 GACAGGCTGGCTTCTTGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTGAC 1380
DB 1321 GACAGGCTGGCTTCTTGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTGAC 1380

QY 1381 TATAACCTCCCGATGAAACCCAGATTTGGGAGCAACAGGTCACTTGTACCTCTGTCT 1440
DB 1381 TATAACCTCCCGATGAAACCCAGATTTGGGAGCAACAGGTCACTTGTACCTCTGTCT 1440

QY 1441 GGGATGGGCGCAGAAATCAGTTTTTCGAGTACACGCTCCAGAAAGAAATACGCTATAACCC 1500
DB 1441 GGGATGGGCGCAGAAATCAGTTTTTCGAGTACACGCTCCAGAAAGAAATACGCTATAACCC 1500

QY 1501 CACAGCTGAGGGCTGCAATGCTGTGGAAGCAGGAATGATACCTTATCATGCTCTC 1560
DB 1501 CACAGCTGAGGGCTGCAATGCTGTGGAAGCAGGAATGATACCTTATCATGCTCTC 1560

QY 1561 TCGGAAGAACTGCCCGCAGAGATCAGAGTTTCATCTTCGAGGAGATGATCTTTATT 1620
DB 1561 TCGGAAGAACTGCCCGCAGAGATCAGAGTTTCATCTTCGAGGAGATGATCTTTATT 1620

QY 1621 CACGAACAGTCCAAGAAATGTTCCAGGCTGCGAGGAGGAGTCCAGTGTGACAGTTTCGTT 1680
DB 1621 CACGAACAGTCCAAGAAATGTTCCAGGCTGCGAGGAGGAGTCCAGTGTGACAGTTTCGTT 1680

QY 1681 CCACTCTTACGAGACTGCAACCACTCGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1740
DB 1681 CCACTCTTACGAGACTGCAACCACTCGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1740

QY 1741 TTATGA 1746
DB 1741 TTATGA 1746

RESULT 2

AB078146

LOCUS

DEFINITION

Homo sapiens

UDP-N-acetyl-alpha-D-galactosamine:polypeptide

N-acetylglucosaminyltransferase 12, complete cds.

AB078146

ACCESSION

VERSION

AB078146 1746 bp mRNA linear PRI 04-SEP-2002
Homo sapiens GALNT12 mRNA for
UDP-N-acetyl-alpha-D-galactosamine:polypeptide
N-acetylglucosaminyltransferase 12, complete cds.
AB078146
AB078146.1 GI:22122073

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 12:45:48 ; Search time 10396.1 Seconds
(without alignments)
11215.070 Million cell updates/sec

Title: US-10-074-527-1
Perfect score: 2850
Sequence: 1 ccgctcggtaccactataa.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba.*
- 2: gb_hg.*
- 3: gb_in.*
- 4: gb_em.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vt.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vt.*
- 30: em_hg_hum.*
- 31: em_hg_inv.*
- 32: em_hg_other.*
- 33: em_hg_mus.*
- 34: em_hg_pln.*
- 35: em_hg_rod.*
- 36: em_hg_mam.*
- 37: em_hg_vrt.*
- 38: em_sy.*
- 39: em_hgo_hum.*
- 40: em_hgo_mus.*
- 41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to
and is derived by analysis of the score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2738.2	96.1	2745	6	AX477705	Sequence
2	2341	82.1	2343	9	AK024865	AK024865 Homo sapi
3	1838.6	64.5	1928	9	BC013945	BC013945 Homo sapi
4	1744.4	61.2	1746	9	AB078146	AB078146 Homo sapi
5	1744.4	61.2	1746	9	HS132365	AB132365 Homo sapi
6	1127	39.5	14930	9	AL136084	AL136084 Human DNA
7	1053.2	37.0	1808	9	AB048801	AB048801 Macaca fa
8	1039	36.5	52737	2	AC007800	AC007800 Homo sapi
9	921.6	32.3	63632	2	AC132018	Continuation (3 of
10	921.6	32.3	110000	2	AC098559	Continuation (3 of
11	911.4	32.0	288118	2	AC129139	Continuation (3 of
12	907.6	31.8	228368	2	AC121727	AC129139 Rattus no
13	906	31.8	288118	2	AC129139	AC121727 Rattus no
14	887.4	31.1	110000	2	AC092373	Continuation (4 of
15	887.4	31.1	166547	2	AC134060	Continuation (2 of
16	873.4	30.6	110000	2	AC098559	Continuation (2 of
17	870.4	30.5	166547	2	AC134060	Continuation (3 of
18	865	30.4	110000	2	AC095991	Continuation (3 of
19	865	30.4	110000	2	AC112324	Continuation (8 of
20	865	30.4	110000	2	AC115635	Continuation (8 of
21	865	30.4	299134	2	AC128783	AC092373 Rattus no
22	853	29.9	110000	2	AC105836	AX197983 Sequence
23	844.8	29.6	94616	2	AC098559	AX208574 Sequence
24	550	19.3	110000	2	AC092373	AC130864 Rattus no
25	527	18.5	554	6	AX197983	AC097343 Rattus no
26	526.4	18.5	544	6	AX208574	U73819 Mus musculu
27	526.2	18.5	328901	2	AC208574	AC101864 Mus muscu
28	522.4	18.3	226623	2	AC097343	BC036390 Homo sapi
29	520.8	18.3	1737	10	MMU73819	AC025034 Homo sapi
30	520.8	18.3	234109	2	AC101864	AX197725 Sequence
31	499.4	17.5	5369	9	BC036390	AX208332 Sequence
32	499.4	17.5	73481	9	AC025034	AX209583 Sequence
33	497	17.4	652	6	AX197725	AX236647 Sequence
34	496.4	17.4	632	6	AX208332	Y08564 Homo sapien
35	496.4	17.4	632	6	AX209583	AC083808 Homo sapi
36	496.2	17.4	1737	6	AR236647	AX330041 Sequence
37	496.2	17.4	1737	9	HSV08564	AX371123 Sequence
38	493.6	17.3	165175	2	AC083808	AX112431 Ciona int
39	409	14.4	421	6	AX330041	AR153422 Sequence
40	409	14.4	421	6	AX335657	AX376296 Sequence
41	363	12.7	483	6	AX371123	AX697278 Sequence
42	300.8	10.6	2384	3	AX112431	
43	285.6	10.0	1920	6	AR153422	
44	285.6	10.0	2575	6	AX376296	
45	285.6	10.0	2575	6	AX697278	

ALIGNMENTS

RESULT 1
AX477705
LOCUS AX477705
DEFINITION Sequence 17 from Patent W00246426.
ACCESSION AX477705
VERSION AX477705.1 GI:22216860
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1
REFERENCE
AUTHORS
Sanjanwala, M.M., Yao, M.G., Au-Young, J., Baughn, M.R., Arvizu, C.,
King, H.Z., Lee, E.A., Ding, L., Hafalia, A.J., Tang, Y.F., Yue, H.,
Tribouley, C.M., Lu, D.A., Lal, P.G., Warren, B.A., Yang, J.,

2745 bp DNA linear PAT 12-AUG-2002

THIS PAGE BLANK (USPTO)

Walia, N.K., Nguyen, D.B., Gandhi, A.R. and Ison, C.H.

Drug metabolizing enzymes

Patent: WO 0246436-A 17 JUN-2002;

Incyte Genomics, Inc. (US)

Location/Qualifiers

source

1..2745

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/note="Incyte ID No: 2860635CB1"

704 a 612 c 737 g 692 t

Query Match

Best Local Similarity 96.1%; Score 2738.2; DB 6; Length 2745;

Matches 2737; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy	81	ATGTGGGGCGGACGGCGGGGCTGCGCGGGAACTGCGCGGGCGGGGAGCG	140
Db	1	ATGTGGGGCGGACGGCGGGGCTGCGCGGGAACTGCGCGGGCGGGGAGCG	60
Qy	141	CTGTTGGTCTCTGCGCTACTGCGTTGGCGGGGCTGGGCTCGGTGCTGGGCGCAG	200
Db	61	CTGTTGGTCTCTGCGCTACTGCGTTGGCGGGGCTGGGCTCGGTGCTGGGCGCAG	120
Qy	201	CGTGGGGCGGGCGGGGCTGCCGAGCGGGACCCCGCGCACCCCGCGCCCGGGCGG	260
Db	121	CGTGGGGCGGGCGGGGCTGCCGAGCGGGACCCCGCGCACCCCGCGCCCGGGCGG	180
Qy	261	CGGAGCGGGTATGCGCGGGCGGGGCTGCGCGGGAAACGCGCTGGGCGGGGCGAG	320
Db	181	CGGAGCGGGTATGCGCGGGCGGGGCTGCGCGGGAAACGCGCTGGGCGGGGCGAG	240
Qy	321	CGGCTGCGGCTCAGCTGCGAGCGGAGGCTGCGGCTGCGAGGAGGAGGCTGCGGCTG	380
Db	241	CGGCTGCGGCTCAGCTGCGAGCGGAGGCTGCGGCTGCGAGGAGGAGGCTGCGGCTG	300
Qy	381	CACAGATTAACATCTACCTGAGCGACCGCATCTCACTGACCGCGCTGCCGAGCGC	440
Db	301	CACAGATTAACATCTACCTGAGCGACCGCATCTCACTGACCGCGCTGCCGAGCGC	360
Qy	441	TGGACCCGCTGCAAGAGAGAAATATGATATGATTAATTTGCCAGAGCATCTGTT	500
Db	361	TGGACCCGCTGCAAGAGAGAAATATGATATGATTAATTTGCCAGAGCATCTGTT	420
Qy	501	ATCATAGCATTTTATAATGAAGCTGGTCAACTCTCTTGGACAGTTTACAGTGTCTT	560
Db	421	ATCATAGCATTTTATAATGAAGCTGGTCAACTCTCTTGGACAGTTTACAGTGTCTT	480
Qy	561	GAGACATCCCGGATATCTGCTAGAAGAGTGATCTTGTGATGACTACAGTGTATGA	620
Db	481	GAGACATCCCGGATATCTGCTAGAAGAGTGATCTTGTGATGACTACAGTGTATGA	540
Qy	621	GAGACATCCCGGATATCTGCTAGAAGAGTGATCTTGTGATGACTACAGTGTATGA	680
Db	541	GAGACATCCCGGATATCTGCTAGAAGAGTGATCTTGTGATGACTACAGTGTATGA	600
Qy	681	CGCGGCAACAAGAGAGAGGCTGCTGCGAGCCGCGCTGCTGGGCGGCTGCGGCGAGG	740
Db	601	CGCGGCAACAAGAGAGAGGCTGCTGCGAGCCGCGCTGCTGGGCGGCTGCGGCGAGG	660
Qy	741	GCGGATGTTCTGACCTTCTGCACTGTCTGATGATGCCAGAGGCTGCGGCGGCGG	800
Db	661	GCGGATGTTCTGACCTTCTGCACTGTCTGATGATGCCAGAGGCTGCGGCGGCGG	720
Qy	801	CTGCTGACAGGATCCATGAAGAGGATGCGGAGTGGTGTGCGCGGCTGATGATGATC	860
Db	721	CTGCTGACAGGATCCATGAAGAGGATGCGGAGTGGTGTGCGCGGCTGATGATGATC	780
Qy	861	GACTGGAAACCTTCAATACCTTGGGAACTCCGGGAGCCCGAGATCGGCGGTTTCGAC	920
Db	781	GACTGGAAACCTTCAATACCTTGGGAACTCCGGGAGCCCGAGATCGGCGGTTTCGAC	840

Qy	921	TGGAGGCTGGTGTACAGTGGCACACAGTTCTCTGAGAGGAGAGATACGGATGCAATCC	980
Db	841	TGGAGGCTGGTGTACAGTGGCACACAGTTCTCTGAGAGGAGAGATACGGATGCAATCC	900
Qy	981	CCCGTCGATGTCATCAGGCTCCAAACAATGGCTGTGGCTGTCTGTGTAGTAAAGAA	1040
Db	901	CCCGTCGATGTCATCAGGCTCCAAACAATGGCTGTGGCTGTCTGTGTAGTAAAGAA	960
Qy	1041	TATTTGAAATATCTGGGCTTATGATACAGGAATGGAGTTTGGGAGGAGAGAAACCTC	1100
Db	961	TATTTGAAATATCTGGGCTTATGATACAGGAATGGAGTTTGGGAGGAGAGAAACCTC	1020
Qy	1101	GAATTTTCTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAACAACACCCATCTCCCAT	1160
Db	1021	GAATTTTCTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAACAACACCCATCTCCCAT	1080
Qy	1161	GTTGGCCATGTTTTCCTCCCAAGCAAGCTCCTACTCCCGCAACAAGCTCTGGCCAAACAGT	1220
Db	1081	GTTGGCCATGTTTTCCTCCCAAGCAAGCTCCTACTCCCGCAACAAGCTCTGGCCAAACAGT	1140
Qy	1221	GTTCTGTCAGCTGCAAGTATGGATGGAATTTTAAAGAGCTCTACTACCATCGCAACCCC	1280
Db	1141	GTTCTGTCAGCTGCAAGTATGGATGGAATTTTAAAGAGCTCTACTACCATCGCAACCCC	1200
Qy	1281	CGTCCCGCTTGGAACTTTTGGGAGTGTGACAGAGAGGAGCAGCTCCGGGACAAAGCTC	1340
Db	1201	CGTCCCGCTTGGAACTTTTGGGAGTGTGACAGAGAGGAGCAGCTCCGGGACAAAGCTC	1260
Qy	1341	CAGTGTAAAGACTTCAAGTGGTCTTTGGAGACTGTGTATCCAGAACTGCAATGTCCTGAG	1400
Db	1261	CAGTGTAAAGACTTCAAGTGGTCTTTGGAGACTGTGTATCCAGAACTGCAATGTCCTGAG	1320
Qy	1401	GACAGCCTGGCTTCTTCCGGATGCTCCAGAACAAAGGACTTAACAGCTACTGCTTTGAC	1460
Db	1321	GACAGCCTGGCTTCTTCCGGATGCTCCAGAACAAAGGACTTAACAGCTACTGCTTTGAC	1380
Qy	1461	TATAACCTCCCGATGAAACCCAGATTTGTTGGACACACAGGTCACTTGTACCTCTGTCAT	1520
Db	1381	TATAACCTCCCGATGAAACCCAGATTTGTTGGACACACAGGTCACTTGTACCTCTGTCAT	1440
Qy	1521	GGATGGGCGCAATCAGTTTTCAGAGTACAGTCCAGAAAGAAATACGCTATACACCC	1580
Db	1441	GGATGGGCGCAATCAGTTTTCAGAGTACAGTCCAGAAAGAAATACGCTATACACCC	1500
Qy	1581	CACAGCCTGAGGCTGCAATGCTGGAAGCAGGAATGATACCTTATCATGATCTC	1640
Db	1501	CACAGCCTGAGGCTGCAATGCTGGAAGCAGGAATGATACCTTATCATGATCTC	1560
Qy	1641	TGCGAAGAAAATGCCCCAGAGAAATCAGAGTTTCATCTTGCAGGAGATGGATCTTTATT	1700
Db	1561	TGCGAAGAAAATGCCCCAGAGAAATCAGAGTTTCATCTTGCAGGAGATGGATCTTTATT	1620
Qy	1701	CACGAACTGCAAGAAATGTTCCAGGCTGCGAGGAGAGTTCGAGTGCAGTTCGTT	1760
Db	1621	CACGAACTGCAAGAAATGTTCCAGGCTGCGAGGAGAGTTCGAGTGCAGTTCGTT	1680
Qy	1761	CCACTCTTACGAGACTGCAACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCCATG	1820
Db	1681	CCACTCTTACGAGACTGCAACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCCATG	1740
Qy	1821	TTATGAGGCTCTGTATCAAGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCC	1880
Db	1741	TTATGAGGCTCTGTATCAAGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCC	1800
Qy	1881	CAACAAGACTTACCTAAGCAGTGCAGAACCCCAACCAAACTAGGCTGCATGCTTTC	1940
Db	1801	CAACAAGACTTACCTAAGCAGTGCAGAACCCCAACCAAACTAGGCTGCATGCTTTC	1860
Qy	1941	AAGAGGCAATCAATTTTCCATTTTGTGAAAGTTGTGTGATTTAGTAAATAATGTGAATA	2000
Db	1861	AAGAGGCAATCAATTTTCCATTTTGTGAAAGTTGTGTGATTTAGTAAATAATGTGAATA	1920
Qy	2001	GCCTTGTACTTATTTTGGAGAACTTTTAAATGTTCCAAATAACCTTATTTTCAAAGGTA	2060

THIS PAGE BLANK (USPTO)

```

Db 1921 GCTTTGACTATTTTGAGACCTTTTAAATGTTCCAAATACCCCTATTTCACAAAGGTA 1980
Qy 2061 ATCGTAAGATGTTAAACCTTGGTATTAGAAAATTAACAACTTATAATATTTTCTATCA 2120
Db 1981 ATCGTAAGATGTTAAACCTTGGTATTAGAAAATTAACAACTTATAATATTTTCTATCA 2040
Qy 2121 ABAWAWAATTTTACAGTCGTCGCTTTTACTCTCATATTAGCAAAAAGATAAAGATTTTAT 2180
Db 2041 AGATGTATATTTTACAGTCGTCGCTTTTACTCTCATATTAGCAAAAAGATAAAGATTTTAT 2100
Qy 2181 TTTGTTATTTACAAAGATTTCCAGGTACGAAGATATCTGATGGGTGAATACAGTTTCA 2240
Db 2101 TTTGTTATTTACAAAGATTTCCAGGTACGAAGATATCTGATGGGTGAATACAGTTTCA 2160
Qy 2241 AGCAACGTACTTTGCAATTAACCTGATACCTCAGCTGCGGGGTTAAAGTTTTCACGTA 2300
Db 2161 AGCAACGTACTTTGCAATTAACCTGATACCTCAGCTGCGGGGTTAAAGTTTTCACGTA 2220
Qy 2301 TAGAGAGACTGTCACCTAGCAATCTGATTTGATTTTATTCAGGTCAATGAGATCTTTAGA 2360
Db 2221 TAGAGAGACTGTCACCTAGCAATCTGATTTGATTTTATTCAGGTCAATGAGATCTTTAGA 2280
Qy 2361 TGTATTTTAAAAAGAAATGCTTTTGGTTATGTTGTTGCTACACAGTTAACTCATAAT 2420
Db 2281 TGTATTTTAAAAAGAAATGCTTTTGGTTATGTTGTTGCTACACAGTTAACTCATAAT 2340
Qy 2421 GTTCATGTACGCAAGAGGACTAACCAAGCTGAAATCTCAGAGAACTTTTCTTTAC 2480
Db 2341 GTTCATGTACGCAAGAGGACTAACCAAGCTGAAATCTCAGAGAACTTTTCTTTAC 2400
Qy 2481 TAAGCTGAGTCAACTTGAGAGGAACTTCTAACAAATGCGGCACTGTAGTGGCTGTTTC 2540
Db 2401 TAAGCTGAGTCAACTTGAGAGGAACTTCTAACAAATGCGGCACTGTAGTGGCTGTTTC 2460
Qy 2541 TACCACATGACTTTTAAAAAGATGTTTATATCAATTTTAAATTTTATGATACGCTAGTTC 2600
Db 2461 TACCACATGACTTTTAAAAAGATGTTTATATCAATTTTAAATTTTATGATACGCTAGTTC 2520
Qy 2601 AGGAGAAATGATTTTCTATATGAATTCCTTTTCAAGTTTGTTCATTATAACAGT 2660
Db 2521 AGGAGAAATGATTTTCTATATGAATTCCTTTTCAAGTTTGTTCATTATAACAGT 2580
Qy 2661 TATTAATTTAAATCAGCGTTAGAGTTTGTGCTGCTGCAACTGCTGTGAAAATTTCTCTGA 2720
Db 2581 TATTAATTTAAATCAGCGTTAGAGTTTGTGCTGCTGCAACTGCTGTGAAAATTTCTCTGA 2640
Qy 2721 GTAAATCTGATTTTGTGAATGATCCAGACCAACCTCAGATTTTGTCAACCTGATTAAGT 2780
Db 2641 GTAAATCTGATTTTGTGAATGATCCAGACCAACCTCAGATTTTGTCAACCTGATTAAGT 2700
Qy 2781 CAATATGAATGATTTAAAGATGTGAGAAACAAAAAATTTTCACTGATTAAGT 2825
Db 2701 CAATATGAATGATTTAAAGATGTGAGAAACAAAAAATTTTCACTGATTAAGT 2745

```

```

RESULT 2
AK024865
LOCUS AK024865 2343 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21212 fis, clone COL00502.
ACCESSION AK024865
VERSION AK024865.1 GI:10437273
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
REFERENCE Kawabata, A., Hiki, J., Kobatake, N., Inagaki, H., Ikema, Y.,
AUTHORS Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,
Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished

```

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

CDS

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 82.1%; Score 2341; DB 9; Length 2343;

Matches 2338; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

```

Qy 508 CATTTTATATCAAGCTGGTCAACTCTCTTCGACAGTTTACAGTCTCTTGAGACAT 567
Db 1 CATTTTATATGAGACCTTGGTCACTCTCTTCGACAGTTTACAGTCTCTTGAGACAT 60
Qy 569 CCCCAGATATCTGCTAGAGAAGTGATCCTTTGATGATGACTACAGTGTATAGAGACAC 627
Db 61 CCCCAGATATCTGCTAGAGAAGTGATCCTTTGATGATGACTACAGTGTATAGAGACAC 120
Qy 628 TGAAGGAGCGCTTGGCCAAATGAGCTTTTCGGAGCTGCCCAAGGTGCGCTGATCCCGCCA 687
Db 121 TGAAGGAGCGCTTGGCCAAATGAGCTTTTCGGAGCTGCCCAAGGTGCGCTGATCCCGCCA 180
Qy 688 ACAAGAGAGAGCGCTGGTGGAGCCCGCTGCTGGGGCGCTCTCGCGGAGGGCGGATG 747
Db 181 ACAAGAGAGAGCGCTGGTGGAGCCCGCTGCTGGGGCGCTCTCGCGGAGGGCGGATG 240
Qy 748 TTCTGACCTTCTTGACATGTCTGAGTGTGCTGAGTGTGCTGAGAGGGTGGCTGGAGCGCTGCTGC 807
Db 241 TTCTGACCTTCTTGACATGTCTGAGTGTGCTGAGTGTGCTGAGAGGGTGGCTGGAGCGCTGCTGC 300
Qy 808 AGAGGATCCATGAGAGAGAGTGGGAGTGGTGTGTCGCCGGTGAATGTGATGATCGACTGGA 867
Db 301 AGAGGATCCATGAGAGAGAGTGGGAGTGGTGTGTCGCCGGTGAATGTGATGATCGACTGGA 360
Qy 868 ACACCTTCGAATACCTGGGAACTCCCGGGAGCCCAAGATCGGCGGTTTTCAGCTGGAGGC 927
Db 361 ACACCTTCGAATACCTGGGAACTCCCGGGAGCCCAAGATCGGCGGTTTTCAGCTGGAGGC 420
Qy 928 TGGTGTTCACCTGGCACACAGTTCCTTGAGAGGAGAGATACGGATGCAATCCCGCGTCG 987

```

2 (bases 1 to 2343)

Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.

Direct Submission

Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

Location/Qualifiers

1..2343

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="COL00502"

/tissue_type="colon"

/clone_lib="COL"

/notes="cloning vector pME18SFL3"

465..1319

/note="unnamed protein product"

/codon_start=1

/protein_id="BAB15027.1"

/db_xref="GI:10437274"

/translation="MQSPDVIRSPTMAGGLFAVSKYFEYLGSDYDTGNEVNGENLE
FSFRWOCGVLETHPCSHVGHVPKQAPYSRNKALANSVRAAEVWMDFFKELYHRN
PRLPEPDVTERKQLDKCKDFKNFLVETYPDLHPVDEPGRPGFGLMKNKGLTDY
CFDYNPBNQVILYCHGMQNOFFEYTSKEIRYNTHQPEGCIAYEAGMDT
LIMHLCETAPENQKFIQEDSLFHEQSKKCVQAAKRESSDSFVPLLRDCTNSDQK
WFFKERM"

673 a 474 c 563 g 633 t

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 12:45:48 ; Search time 10396.1 Seconds
(without alignments)
11215.070 Million cell updates/sec

Title: US-10-074-527-1

Perfect score: 2850

Sequence: 1 ccggctcggtaccactataa.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_red:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2738.2	96.1	2745	6	AX477705 Sequence
2	2341	82.1	2343	9	AK024865 Homo sapi
3	1838.6	64.5	1928	9	BC013945 Homo sapi
4	1744.4	61.2	1746	9	AB078146 Homo sapi
5	1744.4	61.2	1746	9	AB078146 Homo sapi
6	1127	39.5	149930	9	AL1336084 Human DNA
7	1053.2	37.0	1808	9	AB048901 Macaca fa
8	1039	36.5	52737	2	AC007800 Homo sapi
9	921.6	32.3	63632	2	Continuation (5 of
10	921.6	32.3	110000	2	Continuation (3 of
11	911.4	32.0	288118	2	AC129139 Rattus no
12	907.6	31.8	228368	2	AC121727 Rattus no
13	906	31.8	288118	2	AC129139 Rattus no
14	887.4	31.1	110000	2	Continuation (4 of
15	887.4	31.1	166547	2	AC134060 Rattus no
16	873.4	30.6	110000	2	Continuation (2 of
17	870.4	30.5	166547	2	AC098559 Rattus no
18	865	30.4	110000	2	AC095991 Rattus no
19	865	30.4	110000	2	AC112324 Rattus no
20	865	30.4	110000	2	Continuation (3 of
21	865	30.4	299134	2	AC128783 Rattus no
22	853	29.9	110000	2	Continuation (2 of
23	844.8	29.6	94616	2	Continuation (8 of
24	550	19.3	110000	2	AC099237 Rattus no
25	527	18.5	554	6	AX197983 Sequence
26	526.4	18.5	544	6	AX208574 Sequence
27	526.2	18.5	328901	2	AC130864 Rattus no
28	522.4	18.3	226623	2	AC097343 Rattus no
29	520.8	18.3	1737	10	MMU73819 U73819 Mus musculu
30	520.8	18.3	234109	2	AC101864 Mus muscu
31	499.4	17.5	5369	9	BC036390 Homo sapi
32	499.4	17.5	73481	9	AC025034 Homo sapi
33	497	17.4	652	6	AX197725 Sequence
34	496.4	17.4	632	6	AX208332 Sequence
35	496.4	17.4	632	6	AX209583 Sequence
36	496.2	17.4	1737	6	AR236647 Sequence
37	496.2	17.4	1737	9	HSY08564 Y08564 Homo sapien
38	493.6	17.3	165175	2	AC083808 Homo sapi
39	409	14.4	421	6	AX330041 Sequence
40	409	14.4	421	6	AX335657 Sequence
41	363	12.7	483	6	AX371123 Sequence
42	300.8	10.6	2384	3	AK112431 Ciona int
43	285.6	10.0	1920	6	AR153422 Sequence
44	285.6	10.0	2575	6	AX376296 Sequence
45	285.6	10.0	2575	6	AX697278 Sequence

ALIGNMENTS

RESULT 1
AX477705
LOCUS AX477705 2745 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 17 from Patent WO0246426.
ACCESSION AX477705
VERSION AX477705.1 GI:22216860
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1
AUTHORS Sanjanwala,M.M., Yao,M.G., Au-Young,J., Baughn,M.R., Arvizu,C.,
King,H.Z., Lee,E.A., Ding,L., Rafalia,A.J., Tang,Y.F., Yue,H.,
Tribouley,C.M., Lu,D.A., Lai,P.G., Warren,B.A., Yang,J.,

Walia,N.K., Nguyen,D.B., Gandhi,A.R. and Ison,C.H.

Drug metabolizing enzymes
Patent: WO 0246426-A 17 13-JUN-2002;
Incyte Genomics, Inc. (US)

FEATURES

Location/Qualifiers
1..2745
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 2860635CB1"

BASE COUNT 704 a 612 c 737 g 692 t

ORIGIN

Query Match 96.1%; Score 2738.2; DB 6; Length 2745;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 273; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

```
Qy 81 ATGTGGGGGGCGACGGCGCGCGCGCTGCCCGCGGAACTGCGGCGCGCGCGCGCGGAGCG 140
Db 1 ATGTGGGGGGCGACGGCGCGCGCGCTGCCCGCGGAACTGCGGCGCGCGCGCGGAGCG 60

Qy 141 CTGTTGTTGCTCTGGGGCTACTGGCGTTGGCCGGGCTGGGCTGGGCTGGGCGCGCAG 200
Db 61 CTGTTGTTGCTCTGGGGCTACTGGCGTTGGCCGGGCTGGGCTGGGCTGGGCGCGCAG 120

Qy 201 CGTGGGGCGGGCGCGGGCTGCCGAGCCGGAGCCCGCGCGCACCCCGCGCGCGCGCGCGG 260
Db 121 CGTGGGGCGGGCGGGGCTGCCGAGCCGGAGCCCGCGCGCACCCCGCGCGCGCGCGCGG 180

Qy 261 CGCGAGCCGGTCAATGCCCGCGCGCGCGTGGCGCGGAACTGCGGCGCGCGGGCGAG 320
Db 181 CGCGAGCCGGTCAATGCCCGCGCGCGCGTGGCGCGGAACTGCGGCGCGCGGGCGAG 240

Qy 321 CGCGTGGGGTGCAGCTGCAGGGCGAGAGTGGCGCTGCAGGAGAGAGCTGGCGCTG 380
Db 241 CGCGTGGGGTGCAGCTGCAGGGCGAGAGTGGCGCTGCAGGAGAGAGCTGGCGCTG 300

Qy 381 CACCAGATTAACTACTCAGCGCGCGATCTCACTGCGACCGCGCGCGCGCGCGCGCGCG 440
Db 301 CACCAGATTAACTACTCAGCGCGCGATCTCACTGCGACCGCGCGCGCGCGCGCGCGCG 360

Qy 441 TGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCCAGGACATCTGTT 500
Db 361 TGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCCAGGACATCTGTT 420

Qy 501 ATCATAGCATTTTATATAGAGCTGTGTAAGCTCTCTCTCGGACAGTTTACAGTGTCTT 560
Db 421 ATCATAGCATTTTATATAGAGCTGTGTAAGCTCTCTCTCGGACAGTTTACAGTGTCTT 480

Qy 561 GAGACATCCCGGATATCTGCTAGAGAGAGTATCTTGTAGATGACTACAGTGTAGAG 620
Db 481 GAGACATCCCGGATATCTGCTAGAGAGAGTATCTTGTAGATGACTACAGTGTAGAG 540

Qy 621 GAGCACCCTGAAGGAGCGCTTTGGCAATGAGCTTTGGGAGCTGCCCAAGGTCGCTGATC 680
Db 541 GAGCACCCTGAAGGAGCGCTTTGGCAATGAGCTTTGGGAGCTGCCCAAGGTCGCTGATC 600

Qy 681 CGCGCCAAAGAGAGAGGGCTGGTGGAGCCCGGCTGCTGGGGGGCTGCGGGCGAGG 740
Db 601 CGCGCCAAAGAGAGAGGGCTGGTGGAGCCCGGCTGCTGGGGGGCTGCGGGCGAGG 660

Qy 741 GGGATGTTCTGACCTTCCTGGAATGCTGCTGAGTGGCGACGAGAGGGTGGCGAGCGG 800
Db 661 GGGATGTTCTGACCTTCCTGGAATGCTGCTGAGTGGCGACGAGAGGGTGGCGAGCGG 720

Qy 801 CTGCTGCAGAGGATCCATGAGAGAGAGTGCAGTGGTGGTGGCGGCTGATTGATGTGATC 860
Db 721 CTGCTGCAGAGGATCCATGAGAGAGAGTGCAGTGGTGGTGGCGGCTGATTGATGTGATC 780

Qy 861 GACTGGAAACCTTCGAATACCTGGGAACTCCCGGGAGGCCCGAGATCGCGGTTTCGAC 920
Db 781 GACTGGAAACCTTCGAATACCTGGGAACTCCCGGGAGGCCCGAGATCGCGGTTTCGAC 840
```

```
Qy 921 TGGAGGCTGGTGTTCACGTGGCACACAGTTCTTGAGAGGAGAGGATACGGATGCAATCC 980
Db 841 TGGAGGCTGGTGTTCACGTGGCACACAGTTCTTGAGAGGAGAGGATACGGATGCAATCC 900

Qy 981 CCCTGTCATGTCTACAGGTCTCCAAATGGCTGTGGGCTGTTTGTGTGAGTAAGAAA 1040
Db 901 CCCTGTCATGTCTACAGGTCTCCAAATGGCTGTGGGCTGTTTGTGTGAGTAAGAAA 960

Qy 1041 TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAAACCTC 1100
Db 961 TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAAACCTC 1020

Qy 1101 GAATTTTCTTTAGGATCTGGCAGTGTGGTGGGGTTCTGGAACACACACCCATGTCCTCAT 1160
Db 1021 GAATTTTCTTTAGGATCTGGCAGTGTGGTGGGGTTCTGGAACACACACCCATGTCCTCAT 1080

Qy 1161 GTTGGGCATGTTTTCCTCCAAAGCTCCCTACTCCCGCAACAAAGGCTCTGGGCAACAGT 1220
Db 1081 GTTGGGCATGTTTTCCTCCAAAGCTCCCTACTCCCGCAACAAAGGCTCTGGGCAACAGT 1140

Qy 1221 GTTCTGTCAGCTCAAGTATGGATGGAATTTTAAAGAGCTCTACTACCATCGCAACCCC 1280
Db 1141 GTTCTGTCAGCTCAAGTATGGATGGAATTTTAAAGAGCTCTACTACCATCGCAACCCC 1200

Qy 1281 CGTGCCTGGTGGAACTTTTGGGGATGTGACAGAGAGAGAGAGAGCTCCGGGACAGCTC 1340
Db 1201 CGTGCCTGGTGGAACTTTTGGGGATGTGACAGAGAGAGAGAGAGCTCCGGGACAGCTC 1260

Qy 1341 CAGTGTAAAGACTTCAAGTGTCTTGGGAGCTGTGTATCCAGAACTGCTGTCCTGAG 1400
Db 1261 CAGTGTAAAGACTTCAAGTGTCTTGGGAGCTGTGTATCCAGAACTGCTGTCCTGAG 1320

Qy 1401 GACAGCTGGCTTCTTCGGGATGCTCCAGAACAAAGAGCTAACAGACTACTGCTTTGAC 1460
Db 1321 GACAGCTGGCTTCTTCGGGATGCTCCAGAACAAAGAGCTAACAGACTACTGCTTTGAC 1380

Qy 1461 TATAACCTCCCGATGAAACACAGATTGTGGGACACAGAGTCACTTGTACCTGTCAT 1520
Db 1381 TATAACCTCCCGATGAAACACAGATTGTGGGACACAGAGTCACTTGTACCTGTCAT 1440

Qy 1521 GGGATGGGCGCAGAACTCAGTTTTCAGTATACCTCCAGAAAGAAATACGCTATACACC 1580
Db 1441 GGGATGGGCGCAGAACTCAGTTTTCAGTATACCTCCAGAAAGAAATACGCTATACACC 1500

Qy 1581 CACCGCTGAGGCGTGCATGCTGTGGAGAGAGAGATGATACCTTATCATGTCATCTC 1640
Db 1501 CACCGCTGAGGCGTGCATGCTGTGGAGAGAGAGATGATACCTTATCATGTCATCTC 1560

Qy 1641 TGGGAGAAACTCTCCCGCAGAGAACTCAGAACTTCACTTTCAGGAGGATGGATCTTTATT 1700
Db 1561 TGGGAGAAACTCTCCCGCAGAGAACTCAGAACTTCACTTTCAGGAGGATGGATCTTTATT 1620

Qy 1701 CAGGAACTGTCAGGAAATGTGTCCAGGCTGCGAGGAGAGAGTGCAGTGCAGTTTCGTT 1760
Db 1621 CAGGAACTGTCAGGAAATGTGTCCAGGCTGCGAGGAGAGAGTGCAGTGCAGTTTCGTT 1680

Qy 1761 CCCTCTTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCCTTCAAGAGCGCATG 1820
Db 1681 CCCTCTTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCCTTCAAGAGCGCATG 1740

Qy 1821 TTATGAAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCC 1880
Db 1741 TTATGAAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCC 1800

Qy 1881 CAACAAAGACTTACGTAAAGTGAACAGAACCCCAACAACTAGGCTGATGCTTTG 1940
Db 1801 CAACAAAGACTTACGTAAAGTGAACAGAACCCCAACAACTAGGCTGATGCTTTG 1860

Qy 1941 AAGAGCAATCATTTTGCCTATTTGTGAAGTGTGTGGATTTAGTAAATGTAATAA 2000
Db 1861 AAGAGCAATCATTTTGCCTATTTGTGAAGTGTGTGGATTTAGTAAATGTAATAA 1920

Qy 2001 GCTTTGTACTTATTTTGGAGAACTTTTTTAAATGTTTCCAAATAACCTTATTTTCAAGGGTA 2060
```

```
Db 1921 GCTTGTACTATTTTGGAGAACCTTTTAAAGTTCCTAAATACCTATTTTCAAGGGTA 1980
Qy 2061 ATCGTAAGATGTTAAACCTTTGGTATTTAGAAAATTTAAACCTTTATATTTTCTATCA 2120
Db 1981 ATCGTAAGATGTTAAACCTTTGGTATTTAGAAAATTTAAACCTTTATATTTTCTATCA 2040
Qy 2121 ABAWRWAWATTTTACAGTCGTGCTTTTACTCTCTCATTTAGCAAAAAGATAAAGATTTTAT 2180
Db 2041 AGATGTATATTTTACAGTCGTGCTTTTACTCTCATTTAGCAAAAAGATAAAGATTTTAT 2100
Qy 2181 TTTGGTATTTTACAAGAAATCCAGGTAGCAAGATATCTGATGGTGGTGAATCAGGTTCA 2240
Db 2101 TTTGGTATTTTACAAGAAATCCAGGTAGCAAGATATCTGATGGTGGTGAATCAGGTTCA 2160
Qy 2241 AGCAACGTACTTTGCATTAACACTGATATACCTCAGCTGCGGGGTTAAAGTTTCCCAAGTA 2300
Db 2161 AGCAACGTACTTTGCATTAACACTGATATACCTCAGCTGCGGGGTTAAAGTTTCCCAAGTA 2220
Qy 2301 TAGAGAGACTGTCACCTAGGAACATTTGATTTATTTAGGTCATTTAGATCTTCTAGA 2360
Db 2221 TAGAGAGACTGTCACCTAGGAACATTTGATTTATTTAGGTCATTTAGATCTTCTAGA 2280
Qy 2361 TGTATTTTAAAGAAAGATGCTTTTGGTATGTTGCTACACAGTTAAACACTCCATAAT 2420
Db 2281 TGTATTTTAAAGAAAGATGCTTTTGGTATGTTGCTACACAGTTAAACACTCCATAAT 2340
Qy 2421 GTTCATGTCAGCCAAAGAGACTAAACCAAGCTGAAATCTCAGAGAACAAATTTGCTTTAC 2480
Db 2341 GTTCATGTCAGCCAAAGAGACTAAACCAAGCTGAAATCTCAGAGAACAAATTTGCTTTAC 2400
Qy 2481 TAAGCTGAGTCAACTGAGAGAGAACTTCTTAAAGTCCGACCTGATGCTGCTGCTC 2540
Db 2401 TAAGCTGAGTCAACTGAGAGAGAACTTCTTAAAGTCCGACCTGATGCTGCTGCTC 2460
Qy 2541 TACCACATGACCTTTAAACACATGTTATATCATTTTAAATTTTATGATACGCTAGTGC 2600
Db 2461 TACCACATGACCTTTAAACACATGTTATATCATTTTAAATTTTATGATACGCTAGTGC 2520
Qy 2601 AGGGAGAAATGTAATGTTCTATATGAAATTCCTTTTCAAGTTGTTTCATTAATAACAGT 2660
Db 2521 AGGGAGAAATGTAATGTTCTATATGAAATTCCTTTTCAAGTTGTTTCATTAATAACAGT 2580
Qy 2661 TATTAATTTAAATCAGGTTAGAGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2720
Db 2581 TATTAATTTAAATCAGGTTAGAGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2640
Qy 2721 GTAAATCTGATTTGTGAATGATCCAGACCAACCTGAGATTTTGTCAACCTGATTAAGT 2780
Db 2641 GTAAATCTGATTTGTGAATGATCCAGACCAACCTGAGATTTTGTCAACCTGATTAAGT 2700
Qy 2781 CAATATGAATGATTTAAAGAGATGTGAGAACCAAAAAA 2825
Db 2701 CAATATGAATGATTTAAAGAGATGTGAGAACCAAAAAA 2745

RESULT 2
AK024865
LOCUS Homo sapiens cDNA: FLJ21212 fis, clone COL00502. PRI 29-SEP-2000
DEFINITION AK024865
ACCESSION AK024865
VERSION AK024865.1 GI:10437273
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (nt)
Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y.,
Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,
Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.
NEDO human cDNA sequencing project
Unpublished
```

REFERENCE
AUTHORSTITLE
JOURNAL

COMMENT

FEATURES
source

CDS

BASE COUNT
ORIGIN

Query Match 82.1%; Score 2341; DB 9; Length 2343;
Best Local Similarity 99.8%; Pred.No. 0;
Matches 2338; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 508 CATTTTATTAATGAAGCCTGGTCAACTCTCTCGGACAGTTTACAGTGTCTTTGAGACAT 567
Db 1 CATTTTATTAATGAAGCCTGGTCAACTCTCTCGGACAGTTTACAGTGTCTTTGAGACAT 60
Qy 568 CCCCGGATATCTGTCTAGAGAAGTGTATCTTGTGTAGATGACTACAGTGTATAGAGACCC 627
Db 61 CCCCGGATATCTGTCTAGAGAAGTGTATCTTGTGTAGATGACTACAGTGTATAGAGACCC 120
Qy 628 TGAAGGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGGTGCGCTGATCCGCGCCA 687
Db 121 TGAAGGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGGTGCGCTGATCCGCGCCA 180
Qy 688 ACAAGAGAGAGCGCTTGGTCCGAGCGCGCTGTGGGGCGCTCTGGCGGAGGGCGGATG 747
Db 181 ACAAGAGAGAGCGCTTGGTCCGAGCGCGCTGTGGGGCGCTCTGGCGGAGGGCGGATG 240
Qy 748 TTCTGACCTTCTGTGACTGTCTAGTGTAGTGTGCAACGAAAGGTGGTGTGAGCCCTGCTGC 807
Db 241 TTCTGACCTTCTGTGACTGTCTAGTGTAGTGTGCAACGAAAGGTGGTGTGAGCCCTGCTGC 300
Qy 808 AGAGGATCCATGAGAGAGATCGGAGTGTGTCGCCGTGATTTGATGTATGATCGATCGA 867
Db 301 AGAGGATCCATGAGAGAGATCGGAGTGTGTCGCCGTGATTTGATGTATGATCGATCGA 360
Qy 868 ACACCTTCCAAATACCTGGGGAACCTCGGAGAGCGCCAGATCGCGGTGTTTCTGAGTGG 927
Db 361 ACACCTTCCAAATACCTGGGGAACCTCGGAGAGCGCCAGATCGCGGTGTTTCTGATGAG 420
Qy 928 TGGTGTTCACGTGGGCAACAGTTTCTGTAGAGGAGAGATACGGATGCAATCCCGCTCG 987

2 (bases 1 to 2343)
Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,
Shibahara, T., Tanaka, T. and Nakamura, Y.
Direct Submission
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

Location/Qualifiers

1. 2343
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="COL00502"
/tissue_type="colon"
/clone_lib="COL"
/notes="cloning vector pME18SFL3"
465..1319
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB15027.1"
/db_xref="GI:10437274"

/translation="MQSPVDVIRSPMTMAGGLFAVSKYFEYLGSDYDTGMEVWGGENLE
FGRFVQCQGVLETHPCSHVGVFPKQAPYSRNKALANSVRAAEVWMDPEKELYIHRN
PRALEPFGDTPERKOLRDLCKDFKWFLETVYPELHVDPDRPGFGLMQLGLTDY
CFDYNPDENOIVGHQVILYLCHGMGNQFPEVTSOKEIRYNTHQPEGCTAVAGMDT
LIMHLEETAPENQKRFILQEDSLFHEQSKKCVQAAKESSDSFVPLLRDCTNSDROK
WFFKERML"

BASE COUNT 673 a 474 c 563 g 633 t

Db 721 CTGCTCAGAGGATCCATGAAGAGGAGTCGGCAGTGTGTGCCCCGGTGATTGATGTGATC 780
Qy 861 GACTGGAAACACTTCGGAATACCTGGGGAATCCCGGGAGAGCCCGAGATCGCGGTTTCGAC 920
Db 781 GACTGGAAACACTTCGGAATACCTGGGGAATCCCGGGAGAGCCCGAGATCGCGGTTTCGAC 840
Qy 921 TGGAGGCTGTGTTCAGTGGGACACAGTTCCTGAGAGGAGAGGATACCGGATGCAATCC 980
Db 841 TGGAGGCTGTGTTCAGTGGGACACAGTTCCTGAGAGGAGAGGATACCGGATGCAATCC 900
Qy 981 CCCGTCGATGTCATCAGGTCTCCAAACAATCGCTGGTGGGCTGTTGCTGTGAGTAAGAAA 1040
Db 901 CCCGTCGATGTCATCAGGTCTCCAAACAATCGCTGGTGGGCTGTTGCTGTGAGTAAGAAA 960
Qy 1041 TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTCGGGAGGAGAAAACCTC 1100
Db 961 TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTCGGGAGGAGAAAACCTC 1020
Qy 1101 GAATTTTCCCTTAGGATCTGGCAGTGTGGTGGGTTCTGAAACACACCCATGTTCCCAT 1160
Db 1021 GAATTTTCCCTTAGGATCTGGCAGTGTGGTGGGTTCTGAAACACACCCATGTTCCCAT 1080
Qy 1161 GTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAACAGT 1220
Db 1081 GTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAACAGT 1140
Qy 1221 GTTGTGACGCTGAAGTATGGATGGATGAATTTAAAGAGCTCTACTACCAATCGCAACCCC 1280
Db 1141 GTTGTGACGCTGAAGTATGGATGGATGAATTTAAAGAGCTCTACTACCAATCGCAACCCC 1200
Qy 1281 CGTCCCGCTTGAACTTTTGGGGATGTGACAGAGGAGAGAGCTCCGGGCAAGCTC 1340
Db 1201 CGTCCCGCTTGAACTTTTGGGGATGTGACAGAGGAGAGAGCTCCGGGCAAGCTC 1260
Qy 1341 CAGTGTAAAGACTTCAAGTGTCTTGGAGACTGTGATCCAGACTGATGCTGCTGAG 1400
Db 1261 CAGTGTAAAGACTTCAAGTGTCTTGGAGACTGTGATCCAGACTGATGCTGCTGAG 1320
Qy 1401 GACAGGCTGGCTTCTTGGGGATGCTCCAGAAACAAAGGACTAAGAGACTACTGCTTTGAC 1460
Db 1321 GACAGGCTGGCTTCTTGGGGATGCTCCAGAAACAAAGGACTAAGAGACTACTGCTTTGAC 1380
Qy 1461 TATAACCTCCCGATGAAACCAAGATGTGGGACACAGGTCATTTCTGTACCTCTGTCTAT 1520
Db 1381 TATAACCTCCCGATGAAACCAAGATGTGGGACACAGGTCATTTCTGTACCTCTGTCTAT 1440
Qy 1521 GGGATGGCCAGAAATCAGTTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATACACC 1580
Db 1441 GGGATGGCCAGAAATCAGTTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATACACC 1500
Qy 1581 CACAGCCTGAGGCTGCATTTGCTGGAAGAGGAGGATGATACCTTTATCATGCTATCTC 1640
Db 1501 CACAGCCTGAGGCTGCATTTGCTGGAAGAGGAGGATGATACCTTTATCATGCTATCTC 1560
Qy 1641 TGGGAAGAACTGCCCCAGAGAAATCAGAGTTCACTTTGAGAGGATGATCTTTATTT 1700
Db 1561 TGGGAAGAACTGCCCCAGAGAAATCAGAGTTCACTTTGAGAGGATGATCTTTATTT 1620
Qy 1701 CAGGAACGTCAGAAATGTGTCAGGCTCCAGGAGAGGATGAGTGCAGTTCGTT 1760
Db 1621 CAGGAACGTCAGAAATGTGTCAGGCTCCAGGAGAGGATGAGTGCAGTTCGTT 1680
Qy 1761 CCACCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCCTCAAAGAGCGCATG 1820
Db 1681 CCACCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCCTCAAAGAGCGCATG 1740
Qy 1821 TTATGA 1826
Db 1741 TTATGA 1746

RESULT 5
HSA132365

LOCUS HSA132365 1746 bp mRNA linear PRI 27-NOV-2002
DEFINITION Homo sapiens mRNA for UDP-GalNAC-transferase 12 (GALNT12 gene).
ACCESSION AJ132365
VERSION AJ132365.2 GI:25815115
KEYWORDS GALNT12 gene; UDP-GalNAC-transferase 12.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1746)
AUTHORS Bennett E.P.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-1999) Bennett E.P., Glycobiology Group - 24.5.33,
School of dentistry, Noerre Alle', 20, Copenhagen 2200N, DENMARK
REMARK revised by author [25-NOV-2002]
FEATURES
Location/Qualifiers
1..1746
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..1746
/gene="GALNT12"
1..1746
/gene="GALNT12"
/codon_start=1
/product="UDP-GalNAC-transferase 12"
/protein_id="CAC80100.2"
/db_xref="GI:25815116"
/translation="MWGTRARRRCPRRLRRGREALLVLLALLALAGLGSVLRQNRAG
AGAAFGPPRTPPRRPEVPPRPVPPANALGARGAVALRQLQGEELRLOESVRLHQ
INYSLSRISHRLPERWNPCKKKDYDNLPTSVIAFYNEAWSLTLRTVSVL
ETSPDILLEVILDVDSREHLKERLANELSGLPKVRILIRANKRGLVARLIGASA
AGDVLTFDJDCEHGWLEPLQRIHEESAIVACPVIDVDWNTFEYLGNSGEPOI
CGDFDLVPTVHTVPERERIRMQSPVDVIRSPTMAGGLEFAVSKYFEYLGSDYTGHEV
NGENLEFSFRIWQCGVULETHPCSHVGHVFPKQAPYSRNKALANSVRAEAVWMDFFK
EUYHRNPRARLEPPDQVTERKOLDKQCKDFKMPLETFYVPELHVPERPFGFMLO
NKGLTDYCFDYNPNPDQNIQHVILYLCGNGQNGQFYFTSQKEIRYNTHOPEGCTA
VEAGMDTLIMHLCEETAPEKQFIKLDGSLFHEQSKCKVQAARKSSDSFVPLLRDC
TNSDQKWFKKERML"

BASE COUNT 382 a 449 c 549 g 366 t
ORIGIN

Query Match 61.2%; Score 1744.4; DB 9; Length 1746;
Best Local Similarity 99.9%; Pred. No. 3.8e-275;
Matches 1745; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 81 ATGTGGGGCGGCACCGCGCGGGCGGCTGCCCGCGGGAACCTCGCGCGCGCGCGGAGGCG 140
Db 1 ATGTGGGGCGGCACCGCGCGGGCGGCTGCCCGCGGGAACCTCGCGCGCGCGCGGAGGCG 60
Qy 141 CTGTTGGTGCTCTCTGGCGGTACTTGGCGTTGGCGGGCTGGGCTCGGTCTGGGGCGCAG 200
Db 61 CTGTTGGTGCTCTCTGGCGGTACTTGGCGTTGGCGGGCTGGGCTCGGTCTGGGGCGCAG 120
Qy 201 CGTGGGGCGGGGCGGGGCTGCCGAGCGGGACCCCGCGCGACCCCGCGCGCGCGCGCGG 260
Db 121 CGTGGGGCGGGGCGGGGCTGCCGAGCGGGACCCCGCGCGACCCCGCGCGCGCGCGG 180
Qy 261 CGCGAGCCCGGTATGCGCGGGCGCGCGGTGCGCGGAAACGCGCTGGCGCGCGGGCGGAG 320
Db 181 CGCGAGCCCGGTATGCGCGGGCGCGCGGTGCGCGGAAACGCGCTGGCGCGCGGGCGGAG 240
Qy 321 GCGGTGCGGCTCAGCTGCAGGCGGAGGAGCTGCGGCTCAGAGGAGAGAGCGTGGCGGCTG 380
Db 241 GCGGTGCGGCTCAGCTGCAGGCGGAGGAGCTGCGGCTCAGAGGAGAGAGCGTGGCGGCTG 300
Qy 381 CACCGATTAAACATCTTACCTCAGCGACCCGATCTCACTGCACCGCGCGCTGCCCGAGCGC 440
Db 301 CACCGATTAAACATCTTACCTCAGCGACCCGATCTCACTGCACCGCGCGCTGCCCGAGCGC 360
Qy 441 TGGAAACCCGCTGTGCAAGAGAGAAATATGATGATGATGATGATGATGATGATGATGATGAT 500
Db 361 TGGAAACCCGCTGTGCAAGAGAGAAATATGATGATGATGATGATGATGATGATGATGATGAT 420

```
501 ATCATAGCATTTTATATGAGCCTGGTCAACTCTCTTCGGACAGTTTACAGTGTCCCTT 560
Db
421 ATCATAGCATTTTATATGAGCCTGGTCAACTCTCTTCGGACAGTTTACAGTGTCCCTT 480
Qy
561 GAGACATCCCGGATATCTCTGTAGAGAGAGTGTATCTTGTAGATGACTACAGTGTATAGA 620
Db
481 GAGACATCCCGGATATCTCTGTAGAGAGAGTGTATCTTGTAGATGACTACAGTGTATAGA 540
Qy
621 GAGCACCCTGAAGGAGCGCTTGGCCAAATGAGCTTTTCGGGACTGCCCAAGGTGCGCTGATC 680
Db
541 GAGCACCCTGAAGGAGCGCTTGGCCAAATGAGCTTTTCGGGACTGCCCAAGGTGCGCTGATC 600
Qy
681 CGCGCCAAAGAGAGAGGCGCTGTGCGAGCCCGCTGTGCGGCGCTGTGCGGCGAGG 740
Db
601 CGCGCCAAAGAGAGAGGCGCTGTGCGAGCCCGCTGTGCGGCGCTGTGCGGCGAGG 660
Qy
741 GCGAGTGTCTGACCTTCTCTGGAAGTGTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 800
Db
661 GCGAGTGTCTGACCTTCTCTGGAAGTGTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 720
Qy
801 CTGCTGCAGAGGATCCATGAAGAGGAGTTCGCGAGTGTGCGCGGTGATGTGATGATC 860
Db
721 CTGCTGCAGAGGATCCATGAAGAGGAGTTCGCGAGTGTGCGCGGTGATGTGATGATC 780
Qy
861 GACTGGAACACCTTCGAAATACCTGCGGGAACCTCGCGGAGCGCCAGATCGCGGTTTCGAC 920
Db
781 GACTGGAACACCTTCGAAATACCTGCGGGAACCTCGCGGAGCGCCAGATCGCGGTTTCGAC 840
Qy
921 TGGAGGCTGTGTTCACGTGCGCACACAGTTCCTGAGAGGAGGAGGATACGGATGCAATCC 980
Db
841 TGGAGGCTGTGTTCACGTGCGCACACAGTTCCTGAGAGGAGGAGGATACGGATGCAATCC 900
Qy
981 CCCGTCGATGTCTATCAGGTCTCCAAACATGCGTGTGCGGCTGTGCTGTGAGTAAGAAA 1040
Db
901 CCCGTCGATGTCTATCAGGTCTCCAAACATGCGTGTGCGGCTGTGCTGTGAGTAAGAAA 960
Qy
1041 TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTGGGGAGGAGAAACCTC 1100
Db
961 TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTGGGGAGGAGAAACCTC 1020
Qy
1101 GAATTTTCTTTAGGATCTGCGAGTGTGCGGTTCTGAAACACACACCATGTTCCCAT 1160
Db
1021 GAATTTTCTTTAGGATCTGCGAGTGTGCGGTTCTGAAACACACACCATGTTCCCAT 1080
Qy
1161 GTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTGTGGCAACAGT 1220
Db
1081 GTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTGTGGCAACAGT 1140
Qy
1221 GTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTGTGGCAACAGT 1280
Db
1141 GTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTGTGGCAACAGT 1200
Qy
1281 CGTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTGTGGCAACAGT 1340
Db
1201 CGTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTGTGGCAACAGT 1260
Qy
1341 CAGTGTAAAGACTTCAAGTGTGTTCTTGAGACTGTGTATCCAGAACTGCATGTGCTTGG 1400
Db
1261 CAGTGTAAAGACTTCAAGTGTGTTCTTGAGACTGTGTATCCAGAACTGCATGTGCTTGG 1320
Qy
1401 GACAGGCTGCGCTTCTTGGGATGCTCCAGAAACAAAGGACTAACAGACTACTGCTTGGAC 1460
Db
1321 GACAGGCTGCGCTTCTTGGGATGCTCCAGAAACAAAGGACTAACAGACTACTGCTTGGAC 1380
Qy
1461 TATACCTCTCCGATGAAACCAAGTGTGGGACACAGGCTATTCTGTACTCTGTGAT 1520
Db
1381 TATACCTCTCCGATGAAACCAAGTGTGGGACACAGGCTATTCTGTACTCTGTGAT 1440
Qy
1521 GGGATGGGCCAGAAATCAGTTTTTTCAGTACACGTCCCAAGAAAGAAATACGCTATAACACC 1580
Db
1441 GGGATGGGCCAGAAATCAGTTTTTTCAGTACACGTCCCAAGAAAGAAATACGCTATAACACC 1500
```

1581 CACCAGCCTGAGGCTGCATTCCTGTGAAGCAGGAATGGATACCTTTATCATGATCTC 1640
1501 CACCAGCCTGAGGCTGCATTCCTGTGAAGCAGGAATGGATACCTTTATCATGATCTC 1560
1641 TCGAAGAAATCGCCCCAGAGAAATCAGAAGTTTCATCTTCGAGGAGGATGGATCTTTATTT 1700
1561 TCGAAGAAATCGCCCCAGAGAAATCAGAAGTTTCATCTTCGAGGAGGATGGATCTTTATTT 1620
1701 CACGACAGCTCAAGAAATGTCTCCAGGCTCGAGGAGGAGTCCGAGTGCACAGTTTCGTT 1760
1621 CACGACAGCTCAAGAAATGTCTCCAGGCTCGAGGAGGAGTCCGAGTGCACAGTTTCGTT 1680
1761 CCACCTTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1820
1681 CCACCTTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1740
1821 TTATGA 1826
1741 TTATGA 1746

RESULT 6
AL136084
LOCUS
DEFINITION Human DNA sequence from clone RP11-92C4 on chromosome 9q22.33-31.3, complete sequence.
ACCESSION AL136084
VERSION AL136084.11 GI:11557858
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 149930)
Direct Submission
Submitted (14-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Dec 5, 2000 this sequence version replaced gi:11544910.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: En:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr9
RP11-92C4 is from the library RPC1-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
Vektor: pBACe3.6
This sequence is the entire insert of clone RP11-92C4 The true left end of clone RP11-192E23 is at 100457 in this sequence.

FEATURES
source
1..149930
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/map="q22.33-31.3"


```
/note="AluX repeat: matches 1. .293 of consensus"
repeat_region 32390. .32709
/note="L2 repeat: matches 2325. .2675 of consensus"
repeat_region 33291. .33322

Query Match      39.5%; Score 1127; DB 9; Length 149930;
Best Local Similarity 97.4%; Pred. No. 5.8e-174;
Matches 1139; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

Qy 1582 GGAGATGGATCTTTATTTACGAAACAGTCAAGAAATGTCTCAGGCTGGAGGAGGA 1741
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 46535 GTAGATGGATCTTTATTTACGAAACAGTCAAGAAATGTCTCAGGCTGGAGGAGGA 46594
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1742 GTCAGATGACAGTTTCGTTCCACCTTACGAGAGCTGCACCAACTCGGATCATCAGAAATG 1801
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 46595 GTCAGATGACAGTTTCGTTCCACCTTACGAGAGCTGCACCAACTCGGATCATCAGAAATG 46654
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1802 GTTCTTCAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACT 1861
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 46655 GTTCTTCAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACT 46714
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1862 GTGAGCCAGACTCTGCCCAACAAAGACTTTAGCTAAGCAGTGAACGACCCACCAAAA 1921
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 46715 GTGAGCCAGACTCTGCCCAACAAAGACTTTAGCTAAGCAGTGAACGACCCACCAAAA 46774
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1922 ACTAGGCTGCATTGCTTTGAAGAGGCAATCATTTTGGCATTGTGAAAGTTGTCTGGAT 1981
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 46775 ACTAGGCTGCATTGCTTTGAAGAGGCAATCATTTTGGCATTGTGAAAGTTGTCTGGAT 46834
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1982 TTAGTAAAAATGTGAATAAGCTTTGTACTTATTTTGAAGAACTTTTAAATGTTCCTAAAT 2041
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 46835 TTAGTAAAAATGTGAATAAGCTTTGTACTTATTTTGAAGAACTTTTAAATGTTCCTAAAT 46894
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2042 ACCCTATTTTCAAGGGTAATCGTAAGATGTTAACCTTGTATTTAGAAAATTTAAACCC 2101
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 46895 ACCCTATTTTCAAGGGTAATCGTAAGATGTTAACCTTGTATTTAGAAAATTTAAACCC 46954
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2102 TTATAATATTTTCTATCAARAWAWATTTTACAGTCGTGCCCTTTTACTCTCAATTAGCA 2161
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 46955 TTATAATATTTTCTATCAAGATGTATTTTACAGTCGTGCCCTTTTACTCTCAATTAGCA 47014
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2162 AAAAGATAAAGATTTTATTTTGGTATTTTCAAGAAATTTCCAGGTACGAAGATATCTGCA 2221
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 47015 AAAAGATAAAGATTTTATTTTGGTATTTTCAAGAAATTTCCAGGTACGAAGATATCTGCA 47074
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2222 TGGGTGGAATCAGGTTCAACCAACGACTTTGATTAACCTGATAAATACCTCAGCTGGG 2281
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 47075 TGGGTGGAATCAGGTTCAACCAACGACTTTGATTAACCTGATAAATACCTCAGCTGGG 47134
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2282 GGTAAAGTTTCCAGTATAGAGAGCTGTCAGTACGAAATGTTGATTTTATTCAG 2341
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 47135 GGTAAAGTTTCCAGTATAGAGAGCTGTCAGTACGAAATGTTGATTTTATTCAG 47194
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2342 GTCATTGAGATCTTTCTAGATGATTTTAAAAAGAAATGCTTTTGGTATGTTGCTACC 2401
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 47195 GTCATTGAGATCTTTCTAGATGATTTTAAAAAGAAATGCTTTTGGTATGTTGCTACC 47254
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2402 ACAGTTAACCTCCATATGTTTATGTCAGCAAGAGGACTTAACCAAGCTGAATCTC 2461
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 47255 ACAGTTAACCTCCATATGTTTATGTCAGCAAGAGGACTTAACCAAGCTGAATCTC 47314
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2462 AGAGAACAAATTTGCTTACTAAGCTGAGTCAACTTTGAGAGGGAATCTTCAACAATGCGC 2521
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 47315 AGAGAACAAATTTGCTTACTAAGCTGAGTCAACTTTGAGAGGGAATCTTCAACAATGCGC 47374
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2522 ACTGTAGTGGCTGGTTCTACCACTAGTACTTTAAAAACATGTTTATATCATTTTTTAAT 2581
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 47375 ACTGTAGTGGCTGGTTCTACCACTAGTACTTTAAAAACATGTTTATATCATTTTTTAAT 47434
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2582 TTTATGATACGGTAGTGTACGGGAAATGTAATGTTCTATATGAAATTCCTTTTCAAG 2641
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 47435 TTTATGATACGGTAGTGTACGGGAAATGTAATGTTCTATATGAAATTCCTTTTCAAG 47494
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Qy 2642 TTGTTCATTAAATACAGTATTAAATTAATCAGCGTTAGAGTTGTGTCTGCTGCAACT 2701
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 47495 TTGTTCATTAAATACAGTATTAAATTAATCAGCGTTAGAGTTGTGTCTGCTGCAACT 47554
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2702 GCTGTGAAAATTTCTCTGAGTAATTTCTGATTGTTGTAATGATCCAGACCAACCTTGAT 2761
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 47555 GCTGTGAAAATTTCTCTGAGTAATTTCTGATTGTTGTAATGATCCAGACCAACCTTGAT 47614
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2762 TTGTGCAACCTGATTAAAGTCAATATGAATGATTAATAAAGATGTGAGAACAAAAA 2821
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 47615 TTGTGCAACCTGATTAAAGTCAATATGAATGATTAATAAAGATGTGAGAACACTGCTGAG 47674
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2822 AAAAAAATTTTCTGACTGTAACCAAAAAATAA 2850
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 47675 ATCTTTTGTACTGTAACCAAAAAATAA 47703
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
AB048901 1808 bp mRNA linear PRI 20-SEP-2000
LOCUS Macaca fascicularis brain cDNA, clone:QnpA-17439.
DEFINITION
AB048901
ACCESSION
AB048901.1 GI:10241975
VERSION
AB048901.1
KEYWORDS
Macaca fascicularis (crab-eating macaque)
SOURCE
Macaca fascicularis
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE
1 (sites)
AUTHORS Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M.,
Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
TITLE Isolation of full-length cDNA clones from macaque brain cDNA
libraries
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1808)
AUTHORS Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
TITLE Direct Submission
JOURNAL Submitted (14-SEP-2000) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(S-mail:khashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/,
Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
Lab host: TOP10
Vector: pME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACTGTGTG)
R. Site2: DraIII (CACTGTGTG)
Description: 1st strand cDNA was primed with an oligo (dT) primer
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb.The SfiI-digested PCR product was cloned
into distinct draIII sites of pME18S-FL3. XhoI sites just outside
the draIII sites can be used to isolate the cDNA insert. Libraries
were constructed by Sugano et al.(University of Tokyo, Institute of
Medical Science). Custom primer used for sequencing
( 5' end primer [CTTCTGCTCTAAAGCTGG];
3' end primer [CGACCTCGAGCTCGACCA] ).
FEATURES
source
1. .1808
/organism="Macaca fascicularis"
/mol_type="mRNA"
/db_xref="taxon:9541"
/clone="QnpA-17439"
/sex="male"
/tissue_type="brain parietal lobe"
/clone_lib="macaque brain cDNA library QnpA"
/dev_stage="adult"
BASE COUNT 531 a 365 c 383 g 529 t
ORIGIN
Query Match 37.0%; Score 1053.2; DB 9; Length 1808;
Best Local Similarity 84.9%; Pred. No. 3.7e-162;
```


Matches 1300; Conservative 5; Mismatches 68; Indels 159; Gaps 5;	
Qy	1292 GGAACTTTTGGGATGTGACAGAGAGAAAGCAGCTCCGGACACAGCTCCAGTGTAAAGA 1351
Db	421 GGAACTTTTGGGATGTGACAGAGAGAAAGCAGCTCCGGACCAAGCTCCAGTGTAAAGA 480
Qy	1352 CTTCAAGTGGTCTTTGGAGACTGTGTATCCAGAACTGATGTGCTCGAGACAGGCTCTGG 1411
Db	481 CTTCAAGTGGTCTTTGGAGACTGTGTATCCAGAAATGATGTGCTCGAGACAGGCTCTGG 540
Qy	1412 CTTCTCGGATGCTCCAGAAACAAAGCACTAAACAGACTACTGCTTTGACTATACCTCTCC 1471
Db	541 CTTCTCGGATGCTCCAGAAACAAAGGCTGAAAGATTAATGCTTTGACTATACCTCTCC 600
Qy	1472 CGATGAAACACAGATGTGGGACACCAAGGTCACTCTGTCTGTATGATGGATGGGCCA 1531
Db	601 TGATGAAACACAG----- 613
Qy	1532 GAATCAGTCTTTTCAGGTACAGTCCAGAAAGAAATAGCTATTAACACCCACAGCCTGA 1591
Db	614 ----- 613
Qy	1592 GGGCTGCAATGCTGTGGAAGCAGGAATGGATACCTTTATCATGCACTCTCTGCGAAGAAC 1651
Db	614 -----ATCATGCACTCTCTGCGAAGAAC 636
Qy	1652 TGCCCCAGAGAAATCAGAAAGTTCACTTTGCGAGGAGATGGATCTTTATTTACAGCAACAGTC 1711
Db	637 TGCCCCAGAGAAATCAGAAAGTTCACTTTGCGAGGAGATGGTCTTTATTTACAGCAACAGTC 696
Qy	1712 CAAGAAATGTGTCCAGCTGGAGGAAAGAGTGCAGTGCACAGTTTGGTTCACCTCTTTACG 1771
Db	697 CAAGAAATGTGTCCAGCTGGAGGAAAGAGTGCAGTGCACAGTTTGGTTCACCTCTTTACG 756
Qy	1772 AGACTGCACCAACTCGCATCATCAGAAATGCTTCTTCAAAGAGCGCATGTTATCAAGCCT 1831
Db	757 AGACTGCACCAACTCGCATCATCAGAAATGCTTCTTCAAAGAGCGCATGTTATCAAGCCT 816
Qy	1832 CGTGTATCAAGAGGCCATCGAAGGAGACTGTGAGCGCAGGACTCTGCCCAACAAAGACT 1891
Db	817 TGTGTATCAAGAGGCCATCGTAGAGACTGTGAGCGCAGGACTCTGCCCAACAAAGACT 876
Qy	1892 TAGCTAAGCAGTGACCAAGACCCACCAAACTAGGCTGCATTTGCTTTGAAGGCGCAATC 1951
Db	877 TAGCTAAGCAGTGACCAAGAGCCCTCCAAAACCTAGGCTGCTCTCTTTGAGGAGGGAATC 936
Qy	1952 ATTTTGCATTTGTGAAAGTTGTCTGATTTAGTAAATGTAATAAGCTTTGTACTT 2011
Db	937 ATTTTGCATTTGTGAAAGTTGTCTGATTTAGTAAATGTAATAAGCTTTGTACTT 996
Qy	2012 ATTTTGAGAACTTTTAAATGTTCCAAAATACCTATTTTCAAAGGGTAAATCGTAAGATG 2071
Db	997 ACTTTGAAAATCTTTTAAATGTTCCAAAATACCTATTTTCAAAGGGTAAATCATAAAGATG 1056
Qy	2072 TTAACCTTGGTATTTAGAAAAATTAACCTTATAATTTTTTCTATCAARWRWAWATT 2131
Db	1057 TTAACCTTGGTATTT-AAAAATTAACCTTATAATTTTTTCTATCAAGATGTATATT 1115
Qy	2132 TTACAGTCGTGCTTTTACTCTCATTTAGCAAAAAGATAAGATTTTTATTTGGTATTTA 2191
Db	1116 TTACAGTCAGAGCTTTTACTTTTCTATTTAGC--AAAAGTAAGAATTTTTATTTGGATATTTA 1173
Qy	2192 CAAGAATTTCCAGGTACGAAGATATCTGCATGGGTGAAATACAGTTTCAAGCAACGACT 2251
Db	1174 CAAGAATTTCCAGGTAGAGATATCTGCATGGGTGAAATACAGTTTCAACCAACGACT 1233
Qy	2252 TTGCATTAACGTAAATACCTCAGCTCGGGGTTTAAAGTTTTCAGTATAGAGAGACTG 2311
Db	1234 TTGCATTAACGTAAATATCTCAGCTCGGGGTTTAAAGTTTTCAGTATAGAGAGACTG 1293
Qy	2312 TCACTAGAACATGTATTTGATTTATTTTCAAGTCAATGAGATCTTCTAGATGATTTTAAA 2371
Db	1294 ACACTAGGAACACTGTATTGGTTTATTTTCAAGTCAATGAGATCTTCTAGATGATTTTAA 1353

Qy	2372 AAGAATGCTTTTGGTATGTGTTGCTACACAGTTAAACACTCCATATGTTCTATGTCAG 2431
Db	1354 AAGAATGCTTTTGGTATGTGTTGCTACACAGTTAAACACTCCATATGTTCTATGTCAG 1413
Qy	2432 CCAAAGGAGCTAACCAAGCTGAATCTCAGAGACAATTTGCTTTACTAAGCTGAGTC 2491
Db	1414 CCAAAGGAGCTAATCAAGCTGAATCTCAAGACAATTTGCTTTACTAAGCTAAGTA 1473
Qy	2492 AACTTGAGAGC-GAACTTTCTAACAATGCCGCACTGTAGTGTGGCTGGTT----- 2539
Db	1474 AACTTGAGAGCGAATCTTCTAAACATGCCCTCACTATAGTGTGGTGGTTGCTTTAGCAA 1533
Qy	2540 CTACCACTATGACTTTAAAAACATGTTTATATCATTTTAAATTTTATGATACGGTAGTGT 2599
Db	1534 CTACCACTATGACTTTAAAAACATGTTTATACCATTTTAAATTTTATCAGACAGTAGTGT 1593
Qy	2600 CAGGAGAAATGAATGTTCTATATGAAATCTCTTTTCAAGTTTGTTCATTAATACAG 2659
Db	1594 CAGGAGAAATGAATGTTCTCATATGAATTTCTTTTACGAGTTTGTTCATTAATACAT 1653
Qy	2660 TTATTAATTTAAATCAGCGTTAGAGTTTGTGCTGCAACTGCTGTGAAATTTCTCTG 2719
Db	1654 TTAGTAATTTAAATCAGCGTTGAGTTTGTGCTGCCCAACTGCTGTGAAATTTCTCTG 1713
Qy	2720 AGTAATTTCTGATTTGTGAATGATCCAGACCAACCCCTGAGATTTTGTCAACCTGATTAAG 2779
Db	1714 AGTAATTTCTGATTTGTGAATGATCCAGACATCAACCCCTGAGATTTTGTAAACCTGATTAAG 1773
Qy	2780 TCAATATGAATGATTTAAAAAGATGTGAGAACA 2811
Db	1774 TCAATATGAATGATTTAAAAAGATGTGAGAACA 1805

RESULT 8

AC007800/c	52737 bp	DNA	linear	HTG 30-JUN-2000
LOCUS	Homo sapiens chromosome 9 clone RP11-105G5 map 9, *** SEQUENCING IN			
DEFINITION	PROGRESS ***, 5 unordered pieces.			
AC007800	AC007800.4	GI:8844153		
VERSION	HTG; HTGS_PHASE1.			
KEYWORDS	Homo sapiens			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 52737)			
AUTHORS	Birren, B., Linton, J., Nussbaum, C. and Lander, E.			
TITLE	Homo sapiens chromosome 9, clone RP11-105G5			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 52737)			
AUTHORS	Birren, B., Linton, J., Nussbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArillano, K., Depayre, E., Devon, K., Dewar, K., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A., Lehotsky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Strange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.			
TITLE	Direct Submission			
JOURNAL	Submitted (12-JUN-1999) Whitehead Institute/MIT Center for Genome Research, 370 Charles Street, Cambridge, MA 02141, USA			
COMMENT	On Jun 30, 2000 this sequence version replaced gi:8705145. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html			

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIER
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L871
 Center clone name: 105_G_5

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 2190: contig of 2190 bp in length
 * 2191 2290: gap of 100 bp
 * 2291 6766: contig of 4476 bp in length
 * 6767 6866: gap of 100 bp
 * 6867 17075: contig of 10209 bp in length
 * 17076 17175: gap of 100 bp
 * 17176 23359: contig of 6184 bp in length
 * 23360 23459: gap of 100 bp
 * 23460 52737: contig of 29278 bp in length.

FEATURES

Location/Qualifiers
 1..52737
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="9"
 /map="9"
 /clone="RP11-105G5"
 /clone_lib="RPC1-11 Human Male BAC"
 BASE COUNT 14823 a 11875 c 11514 g 13712 t 813 others
 ORIGIN

Query Match 36.5%; Score 1039; DB 2; Length 52737;
 Best Local Similarity 96.8%; Pred. No. 1.2e-159;
 Matches 1139; Conservative 5; Mismatches 25; Indels 8; Gaps 8;
 QY 1682 GGAGGATGGATCTTTATTTACGAAACAGTCCAGAAATGTCTCCAGGCTCGAGGAAGGA 1741
 DB 24939 GTAGATGGATCTTTATTTACGAAACAGTCCAGAAATGTCTCCAGGCTCGAGGAAGGA 24880
 QY 1742 GTCAGTGACAGTTTCTGTTCCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATG 1801
 DB 24879 GTCAGTGACAGTTTCTGTTCCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATG 24820
 QY 1802 GTTCTTCAAGAGCGCATGTATCAAGCCTGCTATCAAGGAGCCCATCGAAGGAGACT 1861
 DB 24819 GTTCTTCAAGAGCGCATGTATCAAGCCTGCTATCAAGGAGCCCATCGAAGGAGACT 24760
 QY 1862 GTGGAGCAGGACTCTGCCCAACAAAGACTTAGCTAAGCAGTGACC-AGAACCACCAAAA 1920
 DB 24759 GTGGAGCAGGACTCTGCCCAACAAAGACTTAGCTAAGCAGTGACCAGGAGCCACCAAAA 24700
 QY 1921 AACTAGGCTGCATTGCTTTGAAGAGGCAATCATTTGCCATTTGTGAAAGTTGTGGA 1980
 DB 24699 AACTAGGCTGCATTGCTTTGAAGAGGCAATCATTTGCCATTTGTGAAAGTTGTGGA 24640
 QY 1981 TTTAGTAAATGTGAATAGCTTTGTACTTATTTTGGAGACTTTTAAATGTTCCAAAA 2040
 DB 24639 TTTAGTAAATGTGAATAGCTTTGTACTTATTTTGGAGACTTTTAAATGTTCCAAAA 24580
 QY 2041 TACCCTATTTTCAAGGGTAAATCGTAAGATGTTAAACCTCTGGTA-TTTAGAAAAATTAAAA 2099
 DB 24579 TACCCTATTTTCAAGGGTAAATCGTAAGATGTTAAACCTCTGGTANTTTAGAAAAATTAAAA 24520
 QY 2100 CCTTATAATATTTTCTATCAARAWRAWATTTT-ACAGTCGTGCTTTTACTCTCATTA 2158
 DB 24519 CCTTATAATATTTTCTATCAAGATGTATATTTTNNACAGTCGTGCTTTTACTCTCATTA 24460

QY 2159 GCAAAAAAGATAAAGATTTTATTTTGGTATTTTACA-AGAATTTCCAGGTACG-AAGATAT 2216
 DB 24459 GCAAAAAAGATAAAGATTTTATTTTGGTATTTTACANAGATTTCCAGGTACGNAAGATAT 24400
 QY 2217 CTCATGGGTGGAATCAGGTTCAAGCAA-CGTAATTTTGGTATTTTAACTGATAAATACCTCAG 2275
 DB 24399 CTCATGGGTGGAATCAGGTTCAAGCAANCGTACTTTGCATTTAACTGATAAATACCTCAG 24340
 QY 2276 CTGCGGGTTAAAGTTTTC-CCAGTATAGAGACTGTCTACTAGGACATTTGTATTGATT 2334
 DB 24339 CTGCGGGTTAAAGTTTTCNCCAGTATAGAGACTGTCTACTAGGACATTTGTATTGATT 24280
 QY 2335 TATTCAGGTCATTGAGATCTTCTAGATG-TATTTTAAAAAGAAATGCTTTTGGTTATGTG 2393
 DB 24279 TATTCAGGTCATTGAGATCTTCTAGATGNTATTTTAAAAAGAAATGCTTTTGGTTATGTG 24220
 QY 2394 TTCTACCAAGTTAAACATCTCCATAATTTTCATGTGTCAGCCAAAGAGACTTAACCAAGCT 2453
 DB 24219 TTCTACCAAGTTAAACATCTCCATAATTTTCATGTGTCAGCCAAAGAGACTTAACCAAGCT 24160
 QY 2454 GAAATCTCAGAGAAACAAATTTTCTTAACTAGCTGAGTCACTTTGAGAGCGAACTTCTAAC 2513
 DB 24159 GAAATCTCAGAGAAACAAATTTTCTTAACTAGCTGAGTCACTTTGAGAGCGAACTTCTAAC 24100
 QY 2514 AATGCCGCACTGTAGTGTGCTGTTCTACCACTATGACTTTTAAACATGTTTATATCAT 2573
 DB 24099 AATGCCGCACTGTAGTGTGCTGTTCTACCACTATGACTTTTAAACATGTTTATATCAT 24040
 QY 2574 TTTTAAATTTTATGATACGCTAGTGTGTCAGGAGAAATGTAATGTTCTATATGAATTCCT 2633
 DB 24039 TTTTAAATTTTATGATACGCTAGTGTGTCAGGAGAAATGTAATGTTCTATATGAATTCCT 23980
 QY 2634 TTTTCAAGTTTGTTCATTATAACAGTTTAAATTTAAATCAGGTTAGAGTTTGTCTG 2693
 DB 23979 TTTTCAAGTTTGTTCATTATAACAGTTTAAATTTAAATCAGGTTAGAGTTTGTCTG 23920
 QY 2694 CTGCAACTGCTGTGAAAAATTTCTCTGAGTAATTTCTGATTAATTTGTAATGATCCAGACCAAC 2753
 DB 23919 CTGCAACTGCTGTGAAAAATTTCTCTGAGTAATTTCTGATTAATTTGTAATGATCCAGACCAAC 23860
 QY 2754 CTTGAGATTTTGTCAACCTGATTAAAGTCAATATGAATGATTAATAAAGATGTGAGAACAA 2813
 DB 23859 CTTGAGATTTTGTCAACCTGATTAAAGTCAATATGAATGATTAATAAAGATGTGAGAACAA 23800
 QY 2814 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 2850
 DB 23799 GACTGCAGATCTTTTGTGTAATTAACCAAAAAAATAA 23763

RESULT 9
 AC132018.4
 WPCOMMENT

Sequence split into 5 fragments LOCUS AC132018 Accession AC132018

Fragment Name	Begin	End
AC132018.0	1	110000
AC132018.1	100001	210000
AC132018.2	200001	310000
AC132018.3	300001	410000
AC132018.4	400001	463632

Continuation (5 of 5) of AC132018 from base 400001 (AC132018 Rattus norvegicus clone CH2).

Query Match 32.3%; Score 921.6; DB 2; Length 63632;
 Best Local Similarity 69.4%; Pred. No. 1.9e-140;
 Matches 1552; Conservative 1; Mismatches 555; Indels 128; Gaps 17;
 QY 468 TATGATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATTAATGAAGCCTGG 527
 DB 44126 TATGATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATTAATGAAGCCTGG 44185
 QY 528 TCAACTCTCTCGGACAGTTTACAGTGTCTTGGAGACATCCCGGATATCTCTAGAA 587
 DB 44186 TCCACACTCTCTCGGACAGTTTACAGTGTCTTGGAGACATCTCCCGTATATCTCTCGAG 44245

Qy	588	GAAGTGATCCTTGTAGATGACTACAGTGATAGAGACACCTGAAGAGAGCCCTTGGCCAAAT	647	Qy	1668	AAATTCAATCTTGCAGGAGGATGGATCTTTATTTTCAAGAACAGTCCCAAGAAATGTGTCAG	1727
Db	44246	GAGGTCAATCTTGTGTAGATGACTACAGCGACAGAGATCACCTGAAGAGAGCCCTTGGTCAA	44304	Db	45316	GAGTTCAATCTACAGGAGGACGCGACGTTAGTCTACAAGCAGAGCGGAAATGTGTGGAG	45375
Qy	648	GAGCTTTCGGGACTTGCCTCAAGGTGCGCTCATCCGCGCCCAACAGAGAGAGGCGCTGGTG	707	Qy	1728	GCTGCGAGGAGAGTGCAGTGACAGTTCGTTCCACTCTTACGAGACTGCACCACTCG	1787
Db	44305	-GAGCTTTCAGAGTGCCTTACAGGTGCGCTCATCCCTCTCTAGCA--AGAGAGGACCTAGTG	44361	Db	45376	GCACAGAGGAGGAGTTCAACTCTGCTTTCGCTCCGAACCTTGCAGAGACTGTAGCACTCA	45435
Qy	708	CGAGCCCGGCTGCTGGGGCGCTCTGCGCGAGGGCGGATGTCTGACCTTCTCGGACTGT	767	Qy	1788	GATCATCAGAAATGGTTCTTCAAGAGCGCATGTATGAAGCCCTCGTGTATCAAGGAGCC	1847
Db	44362	CAAG-CCGGCTACTGGGAGCTTTTGGCGCCAGGGCGAAGTGTCTGAGTTTCTGACTGT	44420	Db	45436	GACAGCCAGAGGTGGTCTTCAAGGAGCAGATGTATGTACGACGCGCTTGTGAGGAAG	45495
Qy	768	CACGTGATGCCACGAGGGTGGTGGAGCCGCTGCTGCAGAGGATCCATGAAGAGGAG	827	Qy	1848	CATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTACGTAAGCAGTGACC	1907
Db	44421	CACGTGACTGTCTAGAGGGTGGTGGAGTCCCTGCTGCAGAGGATCCACGAGAGGAG	44480	Db	45496	GAGCCCAAGCAGCTGTGGCTGTGTGCC-----	45524
Qy	828	TCGGCAGTGTGTGCCCGGTGATTGATGATGACCTGGACCAACCTTCGATACCTGGGG	887	Qy	1908	AGAACCCACCAAAACCTAGCGCTGCATTGCTTTGAAGAGGCAATCATTTTGCATTTGTGA	1967
Db	44481	TCGGCGGTGTGTGCCCGGTATCGATGTCAITGACTGGAAACCTCTGATACCTGGGC	44540	Db	45525	-----CAGTGAGAGGACAAAAGTACGTGGTGTGTCTCTGTGA	45561
Qy	888	AACTCCGGGAGCCCAAGATCGCGGTTTCGACTGGAGGCTGGTGTTCACGTGGCACACA	947	Qy	1968	AGTTCTGTGGATTTAGTAAAAATGTAAATAGCTTTTGTACTTATTTTGTAGAACTTTT	2027
Db	44541	AACTCCGGAGAGCCCAAGATTTGGTGGCTTTGACTGGCGCTAGTATTCACGTGGCAGTG	44600	Db	45562	A---TGTGTGGACTCCACACACATGTGAATATATTTGTAGAGATTTGGAAAACTTTAA	45618
Qy	948	GTTCTGTAGAGGAGAGGATPACGGATGCAATCCCGCGTGCATGTCACTCAGGTCTCCAACA	1007	Qy	2028	AAATGTTCCAAATACCCCTATTTTCAAGGGTAACTCGTAAGATGTAAACCTTTGGTATTT	2087
Db	44601	GTCCCTTGGCGGAGCGGAAGTTGATGCGGACTCCTATTGACATATATCAGGTCTCCAAC	44660	Db	45619	AAATGTTCTCCAGTATCCGTTTCTTAAGGCGCTCTAAGATGTGCTTGTGTTTGTG	45678
Qy	1008	ATGGCTGTGGCTGTGTGGTGTGTGAGTGAAGAAATATTTGAATATCTGGGCTCTTATGAT	1067	Qy	2088	AGAAATTTAAACCTTATAATATTTTCTATCAABAWAWATTTTACAGTCTGCTGCTTT	2147
Db	44661	ATGGCTGTGGGACTGTGTGTGTGTGAGTGAAGATATTTTGAATACCTGGGGTCTTATGAT	44720	Db	45679	GGAGCTTAATGACGA-----ACTGCATTTTACCCTATGCTGCTTT	45720
Qy	1068	ACAGGAATGGAATTTTGGGAGAGAGAAACCTCGAAATTTTCTTTAGGATCTGGCAGTGT	1127	Qy	2148	TACTCTCATTAGCAAAAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTTCCCAAGTA	2207
Db	44721	ACCGAATGGAATCTTGGGAGAGAGAAACCTTGAATTTCTCTTTAGGATCTGGCAGTGT	44780	Db	45721	TATTTCTCATTAGCAGAAAGGGAATATTTTACTTTTGTACTTACAGAGCTTCCCGGGTG	45780
Qy	1128	GTTGGGGTCTCGAAACACACCACTGTTCCTATGTTGGCCATGTTTTCCTCAAGCAAGCT	1187	Qy	2208	CGAAGATATCTGCATGGGTGGAAATCAGGTTCAAGCAACGTAATTTTGCATTAACGATAA	2267
Db	44781	GST--GACTCTGGAACACACACCTTACTCCAGCTGGGCGACGCTCTTCCCTAAGCAAGCT	44838	Db	45781	TTTAGTATATCTGA-----CACAACTGGTAAATACCTCAAAATGTCGACTGTAA	45829
Qy	1188	CCCTACTCCGCAACAGGCTCTGCGCMAAGTGTTCGTGCAGCTGAAGTATCGATGAT	1247	Qy	2268	TACCTCA--GCTGCGGGTTAAAGTTTTCAGTATAGAGAGACTGTCTACTAGAACATTT	2325
Db	44839	CCCTATTCAGCGAGAGGCTCTAGCCACACAGTGTCCGAGCTGCAGAAAGTGTGGACGGAT	44898	Db	45830	TACCTCAATGTCTCAAGTTAGAGTTCTCTCAGGGGGATGAGATGGACACTAGAAAACT	45889
Qy	1248	GAAATTTAAAGAGCTCTACTACCATCGCAACCCCGCTGCGCGCTTGAACCTTTTGGGAT	1307	Qy	2326	GTATTCATTTATTTTCAG-----GTCAATTCAGATCTTCTAGATGTATTTTAAAGAAATGC	2379
Db	44899	GAACTTAAAGAACTCTACCAACCCCGGAAACCCCG--AGCCCGCTGGAAACCTCTGGGAG	44957	Db	45890	GTCTTAATTTTGGTGTGGTGGGAGAGTCTAGATTTTATACAGAAAGAAATGC	45949
Qy	1308	GTGACAGAGAGGAGCAGCTCCGGGCAAGCTCCAGTGTAAAGACTTCAAGTGGTTCTTG	1367	Qy	2380	TTTTTGGTTATGTGTGTGTACCAAGTTAACTC---CATATGTTTCATGTGCGCCAAA	2436
Db	44958	CTGACAGAGAGAAAGAGCTTACAGCCAGCTCCAGTGTGAGAGCTTAAAGCGTTTCTTG	45017	Db	45950	TTTTTGTCTTATGTGTGTGATGCCACAGTCAAACTCCCATCATACGCTCAGACTGACCCAA	46009
Qy	1368	GAGACTGTGTATCCAGACTGCTGCTGAGGACAGGCTGCTTCTCGGATGCTC	1427	Qy	2437	GAGGACTAACCAAGCTGAAATCTCAGAG--AACAAATTCGTTTCTAAGCTGAGTCAACT	2495
Db	45018	GCTACACTGTATCCAGAACTGTATGCTGAGGACAGGCTGTTCTTCGGGAAGCTT	45077	Db	46010	ACAGGCTCACCATCTCTGAAAGCTCAAGAAAGAAATATGTTTGCACAGGCTCTTACAAAC	46069
Qy	1428	CAGAAACAAAGGACTAAGAGACTACTGCTTTGATATATAACCTCCGATGAAACACAGATT	1487	Qy	2496	T---GAGAGCGAACTTCTAACAATCGCCACTGTAGTGTGGCTGGT-----T	2539
Db	45078	GAGAACAGAGGACTACGGGGTACTGCTGGACTATATCTCTCCAGTGAACAAATGTG	45137	Db	46070	TTGCGAGTGTGATTTTCAGCAATGTCTAGTGGAGGATGTCTAGTGACTTAGGAGACT	46129
Qy	1488	GTGGGACACAGGTCAATCTGTACTCTGTATGGGATGGCCAGAAATCAGTTTTCGAG	1547	Qy	2540	CTACCACTATGACTTTTAAACATGTTTATATCATTTTTTAAATTTTTT---ATGATACGGTA	2595
Db	45138	GAAGGCCACAGGCTCTTTGTACCTCTGCGACGGATGGTGCAGAACAGTTTTCGAG	45197	Db	46130	TAACTACTGTGACTGGAAAAAGCATTTATATCATTTCTAGTTTAAATCAACAAATGACA	46189
Qy	1548	TACACGTCCAGAAAGAAATACGCTATACACCCACAGCCTGAGGGCTGCATGTCTGTG	1607	Qy	2596	GTGTGAGGAGAAATGTAATGTTCTATATGAATTCCTTTTCAAGTTTGTTCATTAATA	2655
Db	45198	TGTACATCCGGCAAGAAATACGGTATACACCCCGCAGCCCGCAGGTCTGCTATGAGTGTG	45257	Db	46190	GTTCAGGAGAAACGTAATGTTCTACATGAACCTCAACTTTTAAAGTTTTCCTCTAAATA	46249
Qy	1608	GAAGCAGGAATGATACCTTATCATGCTATCTCTGGAAGAACTGCCCGCAGAGATCAG	1667	Qy	2656	ACAGTTTATTAATTTTAA	2671
Db	45258	GAGGAAGGAGGAT--CCTTATCCCGGATCTCTGGAAGAGATACCGTTCCAGAGAAATCAA	45315	Db	46250	CTATTTGATTTTTTAA	46265

RESULT 10

AC098559 2/c

WPCOMMENT

Sequence split into 8 fragments LOCUS AC098559 Accession AC098559

Fragment Name

Begin End

AC098559_0 1 110000

AC098559_1 100001 210000

AC098559_2 200001 310000

AC098559_3 300001 410000

AC098559_4 400001 510000

AC098559_5 500001 610000

AC098559_6 600001 710000

AC098559_7 700001 794616

Continuation (3 of 8) of AC098559 from base 200001 (AC098559 Rattus norvegicus clone CH2)

Query Match 32.3%; Score 921.6; DB 2; Length 110000;

Best Local Similarity 69.4%; Pred. No. 2.1e-140;

Matches 1552; Conservative 1; Mismatches 555; Indels 128; Gaps 17;

Qy	468	TATGATTATGATAATTTGCCGAGACATCTGTTATCATAGCATTTTATATGAAGCTGG	527
Db	95301	TATGATTATGATAACCTGCGCCAGACATCCGTTGTCTATGATTTTATATGAAGCTGG	95242
Qy	528	TCAACTCTCTTGGCAGATTTACAGTGTCTTGAGACATCCCGGATATCCTGCTAGAA	587
Db	95241	TCACACTCTTTCAGACAGATTTACAGTGTCTTGAGACTTCCCTGATATCCTGCTGAG	95182
Qy	588	GAAGTGATCTCTTGPATGACTACAGTGATAGAGACACTGAAGGAGCCCTTGGCCAAT	647
Db	95181	GAGTCTATCTGTTAGATGACTACAGGACAGAGATCACCTGAAGGAGCCCTGGTCAA	95123
Qy	648	GAGCTTTCCGAGCTGCCCAAGTGGCTGATCCGCCCAACAGAGAGAGGCGCTGGTG	707
Db	95122	-GAGCTTTCACAGCTGCCCTAGTGTGACCTGATCCCTCTAGCA--AGAGAGGACCTAGT	95066
Qy	708	CGAGCCCGCTGCTGGGGGCTGTCGGCGAGGGGCGATGTTCTGACCTTCTCTGGACTGT	767
Db	95065	CAAG-CCGGCTACTGGGAGCTTTGCGCCAGGGCGAGTGCTGAGGTTTCTGGACTGT	95007
Qy	768	CACGTGAGTGCCACGAGGGTGGCTGGAGCCGCTGTCGAGAGGATCCATGAAGAGAG	827
Db	95006	CACGTGACTGTCTAGGGGGTGGCTGGAGTCCCTGCTGAGAGGATCCACGAGAGGAG	94947
Qy	828	TCGCGAGTGTGCTCCCGGTGATGTGATGATGATGATGATGATGATGATGATGATGAT	887
Db	94946	TCGCGCGTGTGTGCTCCCGGTATCATGATGATGATGATGATGATGATGATGATGAT	94887
Qy	888	AACCTCCGGGAGCCACAGATCGGGGCTTTCGACTGGAGGCTGGTGTTCAGCTGGCACCA	947
Db	94886	AACCTCCGGAGAGCCACAGATGGTGGCTTTCGACTGGCGCTAGTATTCAGCTGGCAGGT	94827
Qy	948	GTTCTCAGAGGAGAGAGATACGATGCAATFCCCGCTGATGATGATGATGATGATGATGAT	1007
Db	94826	GTCCCTTGGCGGAGCGGAGTTCGATGCGGACTCCTATTCGACATTCAGCTGCCAAT	94767
Qy	1008	ATGCTGTGGCTGTGTTGCTGCTAGTAAGAAATATTTGAAATCTGGGCTCTTATGAT	1067
Db	94766	ATGCTGTGGGAGCTGTGTTGCTGCTAGTAAGAAATATTTGAAATCTGGGCTCTTATGAT	94707
Qy	1068	ACAGGAATGGAAGTTTGGGAGAGAGAAACCTCGAAATTTCTTTAGGATCTGGCAGTGT	1127
Db	94706	ACCGGAATGGAAGTTTGGGAGAGAGAAACCTTGAATTCCTTTAGGATCTGGCAGTGT	94647
Qy	1128	GGTGGGTTCTGGAACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCT	1187
Db	94646	GGT--GACTCTGGAACACACCCCTACTCCCACTGGGCCACGTCTTCCCTTAAGCAAGCT	94589
Qy	1188	CCCTACTCCCAAGGCTCTGGCCCAACAGTGTCTGTCAGCTGGAATGATGATGAT	1247
Db	94588	CCCTATTACGACGCAAGGCTCTAGCCCAACAGTGTCTGTCAGCTGGAATGATGATGAT	94529
Qy	1248	GAATTTAAGAGCTCTACTACATCGCAACCCCGTGGCCCTTGGGAGCTTTTGGGGAT	1307

Db	94528	GAACCTTAAAGAACTCTACCAACCCCGGAAACCCCC--AGCCCGCTCGAAACCCCTCTGGGGAG	94470
Qy	1308	GTGACAGAGAGAGAGAGAGCTCCGGGACAGCTCCAGTGTAAAGACTTCAAGTGGTCTCTTG	1367
Db	94469	CTGACAGAGAGAGAGAGAGCTTCCAGCAAGCTCCAGTGTGAAGACTTAAAGCGGTCTCTTG	94410
Qy	1368	GAGACTGTGTATCCAGAACTGCATGTGCTTGAGGACAGCCCTGGCTTCTTCGGGATGCTC	1427
Db	94409	GCTACACTGTATCCAGAACTGCATGTGCTTGAGGACAGCCCTGGTCTTCTTCGGGAGCTT	94350
Qy	1428	CAGAAACAAAGGACTAACAGACTACTGCTTTGACTATAAACCTCCCGATGAACCAAGATT	1487
Db	94349	GAGAACAGAGGACTACGGGGTACTCCCTGGGACTATAATCTCCCACTGAAACCAATGTG	94290
Qy	1488	GTGGGACACAGCTCATCTCTGTACCTCTCTGATGGATGGCCAGAACTCAGTTTTTCGAG	1547
Db	94289	GAAGGCCACAGCTCTCTTTGTACCTCTGCCAGGATGGGTGAGAACCAAGTTTTTCGAG	94230
Qy	1548	TACACGTCCAGAAAGAAATAGCTATAACACCCACAGCTGAGGGCTGCATTGCTGTG	1607
Db	94229	TGTACATCCCGCAGAAATACGGTATACACCCCGCAGCCCGAGCTCTGCATAGCAGTG	94170
Qy	1608	GAAGCAGGAATGATACCTTATCATGCTCTCTGCGAAGAAACTGCCCCAGAGAACTCAG	1667
Db	94169	GAGGAAGGGAAGGAT--CCTTATCCCGGATCTCTGGAAGATACCGTTCCAGAGAACTCA	94112
Qy	1668	AGTTTCATCTTCAGGAGGATGATCTTTATTTTCAGAAACAGTCCCAAGAACTGTGCCAG	1727
Db	94111	GAGTTTCATCTTCAGGAGGAGCGACGTTAGTCTTCAAGCAGAGCGGGAATGTGTGGAG	94052
Qy	1728	GCTGCGAGGAAGAGTGCAGTGCACGTTTCGTTTCCACTCTTTACGAGACTGCACCAACTCG	1787
Db	94051	GCCACAGAGAGGAGTTCACCTGCTCTTCTCCGAACTTGGCGAGACTGTAGCAACTCA	93992
Qy	1788	GATCATCAGAAATGTTTCTTCAAAGAGCGCATGTTATGAAGCCCTGTTGATCAAGAGGCC	1847
Db	93991	GACAGCCAGAGGTGTTCTTCAAGGAGCAGATGTCATAGTGCACGGCTTGTGAGGAAG	93932
Qy	1848	CATCGAAGGAGTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAGCTAAGCAGTGAACC	1907
Db	93931	GAGCCACCGCAGCTGTGGGCTGTGTGCC-----	93903
Qy	1908	AGAAACCCACAAAACACTAGGCTGCATTGCTTTGAAGAGGCAATCATTTTCCCATTTGTGA	1967
Db	93902	-----CACTGAGAGGAGCAAAAGTACGTGCTGTCTCATCTGA	93866
Qy	1968	AGTGTGTTGGAATTTAGTAAATGTGAATAAGCTTTTGTACTTATTTTGAAGAACTTTT	2027
Db	93865	A---TGTGTTGGACTCCACAACTGTGAATTTATTTGTACAGATTTGGAAACCTTTAA	93809
Qy	2028	AAATGTTCCAAATACCTTATTTTCAAAGGGTAACTGTAAGATGTTAAACCTTGGTATTT	2087
Db	93808	AAATGTTCCAGTATCCCGTTTCTTAAGGGCGCTCTAAGATGCTGCTTGGTATTTG	93749
Qy	2088	AGAAATTTAAAACCTTATATATTTTCTATCAARAWAWATTTTACAGTCGTCGCTTT	2147
Db	93748	GGAGCTTAAATGACGA-----ACTGCATTTTACCCATGGTGCCTTT	93707
Qy	2148	TACTCTATTTAGCAAAAAGATTAAGATTTTATTTTGGTATTTTACAGAAATTTCCAGGTA	2207
Db	93706	TATTTCTATTAGCAGAAAGGGGAAATATTTTACTTTTGTACAGATTTCCCGGGTG	93647
Qy	2208	CGAAGATATCTGATGGTGGAAATCAGGTTTCAAGCAACGTACTTTTGCATTTAACTGATAA	2267
Db	93646	TTTAGTATATTTGA-----CACAACTGGTAAATACCTCAATGTGCGACTGGTAA	93598
Qy	2268	TACTCTA--GCTCGGGGTTTAAAGTTTTTCCAGTATATAGAGAGACTGTCACTAGGAACATT	2325
Db	93597	TACCTCAATGTCAAGTTAGTGTCTCTCAGGGGATGAGATGGACACTAGAAATACT	93538
Qy	2326	GTATTTGATTTTATTCAG-----GTCAATGAGATCTTCTAGATGATTTTAAAAAGATGC	2379
Db	93537	GTCTTAAATTTTGTGTTGTTGAGTGGGTGGGACAGTCTAGATTTTATACAAAGAAAGATGC	93478

[illegible]

RESULT	11
AC129139	
LOCUS	
DEFINITION	AC129139 289118 bp DNA linear HTG 11-OCT-2002 Rattus norvegicus clone CH230-293I11, *** SEQUENCING IN PROGRESS ***, 14 unordered pieces.
ACCESSION	AC129139
VERSION	AC129139.2 GI:23829088
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE
AUTHORS

1 (bases 1 to 298118)

Muzny, D. Marie., Mettsack, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alebrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, O., Blair, J., Blankenbush, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gregeorgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, J., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Maheshuwa, L., Loulsegged, H., Lozado, R. J., Lu, X., Ma, J., Mahendrabai, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwunodu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S. Paul, H., Perez, A., Perez, L., Pfannkuch, C.,

Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Fuazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajic, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Soreller, R., Sosa, J.,
Stearns, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 288118)
Worley, K.C.
Direct Submission
Submitted (27-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 288118)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 11, 2002 this sequence version replaced gi:21989925.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRCM
Center clone name: CH230-293I11
----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 214989 bases at least Q40
Consensus quality: 221817 bases at least Q30
Consensus quality: 227320 bases at least Q20
Estimated insert size: 252471; sum-of-contigs
Quality coverage: 4x in Q20 bases; sum-of-con-

```

* -----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 5778: contig of 5778 bp in length
* 5779 5878: gap of unknown length
* 5879 16373: contig of 10495 bp in length

```

```

* 16374 16473: gap of unknown length
* 16474 90420: contig of 73947 bp in length
* 90421 90520: gap of unknown length
* 90521 236782: contig of 146262 bp in length
* 236783 236882: gap of unknown length
* 236883 250122: contig of 13240 bp in length
* 250123 250222: gap of unknown length
* 250223 257033: contig of 6811 bp in length
* 257034 257133: gap of unknown length
* 257134 259071: contig of 1938 bp in length
* 259072 259171: gap of unknown length
* 259172 261008: contig of 1837 bp in length
* 261009 261108: gap of unknown length
* 261109 264448: contig of 3340 bp in length
* 264449 264548: gap of unknown length
* 264549 267548: contig of 3000 bp in length
* 267549 271482: contig of 3834 bp in length
* 271483 271582: gap of unknown length
* 271583 275622: contig of 4040 bp in length
* 275623 275722: gap of unknown length
* 275723 286107: contig of 10385 bp in length
* 286108 286207: gap of unknown length
* 286208 288118: contig of 1911 bp in length.
FEATURES
    source
        1..288118
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:101116"
            /clone="CH230-293111"
        1..1287
            /note="wgs_end_extension"
            clone_end:r7"
        1338..5778
            /note="wgs_end_extension"
            clone_end:r7"
        3086..3988
            /note="clone_boundary"
            clone_end:r7"
            site:Mbol
            end_sequence:RXAIQ54TJB"
        5879..7959
            /note="wgs_contig"
        12280..13622
            /note="wgs_contig"
        14019..16373
            /note="wgs_contig"
        16474..19067
            /note="wgs_contig"
        19118..21281
            /note="wgs_contig"
        21332..22964
            /note="wgs_contig"
        28345..31644
            /note="wgs_contig"
        31695..34111
            /note="wgs_contig"
        40505..44913
            /note="wgs_contig"
        44964..47869
            /note="wgs_contig"
        47920..49820
            /note="wgs_contig"
        49871..52672
            /note="wgs_contig"
        52723..56843
            /note="wgs_contig"
        61887..63717
            /note="wgs_contig"
        72165..76565
            /note="wgs_contig"
        78769..86319
            /note="wgs_contig"

```

```

misc_feature 86370..90420
              /note="wgs_contig"
misc_feature 90521..92438
              /note="wgs_contig"
misc_feature 92489..94959
              /note="wgs_contig"
misc_feature 96097..99542
              /note="wgs_contig"
misc_feature 236883..239542

Query Match      32.0%; Score 911.4; DB 2; Length 288118;
Best Local Similarity 68.8%; Pred. No. 1.1e-138;
Matches 1540; Conservative 1; Mismatches 567; Indels 129; Gaps 16;

QY 468 TATGATTATGTAATTTGCCAGGACATCTGTATCATAGCATTTTATATGAAGCTGG 527
DB 83581 TATGATTATGTAATCTGCCCAAGACATCCGTTGTATAGCATTTTATATGAAGCTGG 83640
QY 528 TCAACTCTCTTCGGACAGTTTACAGTCTCTTGAGACATCCCGGATATCCTGCTAGAA 587
DB 83641 TCCACACTCTCTCAGACAGTTTACAGTGTCTCGAGACTTCCCTGATATCCTGCTGGAG 83700
QY 588 GAAGTGATCTTGTAGATGACTACAGTATAGAGACACTGAGAGAGCGCTTGGCAAT 647
DB 83701 GAGGTCAATTCTGTAGATGACTACAGCGACAGAGATCACTGAAGGAGCCCTGGTCAA- 83759
QY 648 GAGCTTTCCGGACTGCCAAGGTGCCCTGTATCCGCGCCCAAGAGAGAGAGCGCTGGTG 707
DB 83760 -GAGCTTTCACAGCTGCCCTAGGTGCACCTGTATCCCTGCTAGCA--AGAGAGGACCTAGTG 83816
QY 708 CGAGCCCGGCTGCTGGGGGCGTCTCGGGGAGGGGGCGATGTTCTGACCTTCTTGGACTGT 767
DB 83817 CAAG-CGGCTACTGGAGCGCTTTGCGGCCAGGGGCGAAGTGTCTGAGTTTCTGGACTGT 83875
QY 768 CACTGTGAGTGCCACAAAGGGTGGCTGGAGCGGCTCTGTCAGAGGATTCATAGAGAGGAG 827
DB 83876 CACTGTGACTGTCAAGGGGTGGCTGGAGTCCCTGTCAGAGGATTCACAGAGAGGAG 83935
QY 828 TCGGCACTGGTGTGCCGGTGATTGATGTGATCTGAGTGGAAACCTTCGAATACCTGGGG 887
DB 83936 TCGGCGGTGGTGTGCCGGTTATCGATGTCATTTGATGGAACACCTCTGAGTACCTGGGC 83995
QY 888 AACTCCGGGAGGCCCCAGATCGCGGTTTCGACTGGAGGCTGGTGTTCAGTGGCCACACA 947
DB 83996 AACTCCGGAGAGGCCCCAGATTGGTGGCTTTCGACTGGCGCTAGTATTCAGTGGCAGGTG 84055
QY 948 GTTCTCTGAGAGGAGAGGATACGGATGCAATCCCGCGTGATGTCATCAGGTCTCCACA 1007
DB 84056 GTCCCTTGGCGGAGCGGAAGTTGATGCGGACTCTCTATTGACATTTATCAGGTCTCCAAC 84115
QY 1008 ATGGCTGGTGGGCTGTTGCTGTGAGTAAGAAATATTTGAATATCTGGGGTCTTATGAT 1067
DB 84116 ATGGCTGGGGGACTGTTGTTGTTGAGTAAGAGATATTTTGAATACCTGGGGTCTTATGAT 84175
QY 1068 ACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAAATTTTCTTTTAGGATCTGGCAGTGT 1127
DB 84176 ACCGGAATGGAAGTCTGGGAGGAGAAACCTTGAATTTCTCTTTTAGGATCTGGCAGTGT 84235
QY 1128 GGTGGGGTCTCGAAAACACACCCATGTTCCCATGTTGGCCATGTTTTCCTCCCAAGCAAGCT 1187
DB 84236 GGT--GACTCTGGAAAACACACCCCTTACTCCACGTTGGGCCACGCTTCCCTTAAGCAAGCT 84293
QY 1188 CCTACTCCCGCAACAAGGCTCTGGCCACAGTGTTCGTGACGTGAAGTATCGATCGAT 1247
DB 84294 CCTATTTCACGCAAGAGGCTCTAGCCACAGTGTCCGAGCTGCAGAAAGTGTGGACCGAT 84353
QY 1248 GAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGGAAACCTTTTGGGGAT 1307
DB 84354 GAATTTAAAGAACTCTTACCAACCCCGGAAACCCCG-AGCCCGCTTGGAAACCTCTGGGGAG 84412
QY 1308 GTGACAGAGAGGAGAGCAGCTCCCGGCAAGCTCCAGGTAAAGACTTCAAGTGTTCTTG 1367
DB 84413 CTGACAGAGAGAAAGAGCTTCAGCCCAAGCTCCAGTGTGAAGACTTTAAAGCGGTTCTTG 84472

```

Qy	1368	GAGACTGTATCAGAACTGCTGCTGAGGACAGGCGCTTCTTCGGGATGCTC	1427
Db	84473	GCTACACTGTACCCAGAACTGCTGCTGAGGACAGGCGCTTCTTCGGGAGCTT	84532
Qy	1428	CAGAAACAGGACTAACAGACTACTGCTTTGACATATAACCTCCCGATGAACACGATT	1487
Db	84533	GAGAAACAGGACTAACCGGGGCTACTGCTGACATATAATCTCCAGTGAACAAATGTG	84592
Qy	1488	GTGGGACACCAAGGTCATTCTGTACCTCTGTCATGGGATGGGCCAGAAATCAGTTTTCGAG	1547
Db	84593	GAAGGCCACCAAGGTCCTTTTGTACCTCTGCGACCGGATGGTCAAGAACGATTTCGAG	84652
Qy	1548	TACAGTCTCCCAAGAAATACGCTATTAACACCCACAGCCTGAGGGCTGATTCGTGTG	1607
Db	84653	TGTACATCCCGGCAAGAAATACGCTATTAACACCCCGCAGCCCAAGGCTGTCATAGCAGTG	84712
Qy	1608	GAAGCAGGAATGGATACCTTTATCATGCTCTGCGAAGAAATCGCCCGCAGAAATCAG	1667
Db	84713	GAGGAAGGGAAGAT--CCTTATCCCGATCTCTGGAAGATACCGTTCCAGAGAAATCAA	84770
Qy	1668	AAGTTATCTTGCAGGAGGATGGATCTTTATTTCCAGAACAGTCCAAAGAAATGTGTCAG	1727
Db	84771	GAGTTTCATCTACAGGAGGACGACGTTAGTCTTACAGCAGAGCGGGAATGTGTCGAG	84830
Qy	1728	GCTGCGAGGAGGAGTGCAGTGCAGTTTCGTTCCACTTTACGAGACTGCACCAACTCG	1787
Db	84831	GCCACAGAGGAGGAGTTCAACTCTGCTGCTGCTCCGAACTTGCAGAGCTGTAGCAACTCA	84890
Qy	1788	GATCATCAGAAATGGTTCTTCAAGAGCGGATGTTATGAGCCTCGTGTATCAAGGAGCC	1847
Db	84891	GACAGCAGAGGTTGCTTCTTCAAGGAGCAGATCATAGTGCAGCGCCTTGTCCGGGAAG	84950
Qy	1848	CATCGAAGGAGACTGTGCGAGCAGGACTCTGCCCAACAAAGACTTAGCTAAGCAGTGACC	1907
Db	84951	GAGCCACCCGAGCTGTGGCTGTGTGCC-----	84979
Qy	1908	AGAACCACCAAAATAGGCTGATGCTTTGGAAGAGGCAATCATTTTCCCATTTGTGA	1967
Db	84980	-----CAGTGAGAGGACAAAGTACGTGGTGTCTGTCATCTGTGA	85016
Qy	1968	AAGTTGTGTTGATTTAGTAAATGTGAATAGCTTTGACTTATTTTTCAGAACTTTT	2027
Db	85017	A-----TGTGTTGGACTCCACAAATGTGAATATATTTGTACAGATTTGGAATCTTAA	85073
Qy	2028	AAATGTTCCAAAATACCTCTATTTTCAAAGGGTAAATCGTAAGATGTAAACCTTGTGATTT	2087
Db	85074	AAATGTTCCAGTATCCCGTTTCTAAGGGCCGCTCTAAGATGCTGCTTGTGATTTG	85133
Qy	2088	AGAAAATTAACCTTATATATATTTTCTATCAARAWRWAWTTTTCAGTCGTCGCTTT	2147
Db	85134	GGAAGCTTAATGCACGA-----ACTGCATTTTACCCATGTCGCTTT	85175
Qy	2148	TACTCTCATAGCAAAAGATAAGATTTTATTTTGGTATTTACAGAAATTCAGGTA	2207
Db	85176	TATTTCTATTAGCAGAAAGGGAATATTTTACTTTTGTACTTACGAAAGCTTCCCGGGT	85235
Qy	2208	CGAAGATATCTGATGGTGGAAATCAGGTTCAAGCAACGTACTTTGCATTAACTGATAA	2267
Db	85236	TTTAGTATACCTGA-----CACAACTGTAATACCTCAATGTGCACTGGTAA	85284
Qy	2268	TACCTCA--GCTCGGGGTTAAAGTTTTCAGTATAGAGAGACTGTCACTAGGAACATT	2325
Db	85285	TACCTCAATGTCAAGTTAGAGTTCTCTCAGGGGATGAGATGGACACTAGAAATACT	85344
Qy	2326	GTATTGATTTATTCAG-----GTCATGTAGATCTTCTAGATGATTTTAAAAAGATGC	2379
Db	85345	GTCTTAATTTCTGTTGTTGAGTGGGTGGGACAGTCTAGATTTTATCAAGAAAGATGC	85404
Qy	2380	TTTTTGGTTATGTTGCTTACCACAGTTAACTC-----CATATGTTTCATGTCAGCCAA	2436
Db	85405	TTTTTGTCTTATGCTGATGCCACAGTCAAACTTCATATACGCTCAGACTGACCCAA	85464
Qy	2437	GAGGACTAACCAAGCTGAAATCTCAGAGAACAAAT-----TGCTTTACTAAGCTGAGTCA	2492
Db	85465	ACAGGCTCACCATCTCTGAAAGCTCAAGAAAGATATGTTGTCCCAAGCTCTAGCAA	85524
Qy	2493	ACTTGAGAGCGAATCTTAACAATCGCGACTGT-----AGTGTGGCTGGT	2538
Db	85525	ACTTGGAGGTGATCATTTTCAGCAATGTCTATGGAGGATGTCTAGTACTTAGGAGGAC	85584
Qy	2539	TCTACCACACTACTTTTAAACATGTTTATATCATTTTAAATTTT-----ATGATACCGT	2594
Db	85585	TTAATCTACTGACTGGAAGAACATTTATATCATTTCTAGTTTAAATCAACAATGAC	85644
Qy	2595	AGTGTGAGGAGAAATGTAATGTTCTATATGAATTCCTTTTCAAGTTTGTTCATTAAT	2654
Db	85645	AGTTCAGGAGAAAGCTAATGTTTCTACATGAACCTCAACTTTTAAAGTTTTCCTAAAT	85704
Qy	2655	AACAGTTATTAAATTTAA 2671	
Db	85705	ACTATTGATGTTTAA 85721	
RESULT 12			
AC121727			
LOCUS			
DEFINITION			
Rattus norvegicus clone CH230-326G7, *** SEQUENCING IN PROGRESS			
***, 13 unordered pieces.			
AC121727			
AC121727.3 GI:23322328			
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.			
Rattus norvegicus (Norway rat)			
Rattus norvegicus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
Rattus.			
1 (bases 1 to 228368)			
REFERENCE			
AUTHORS			
Muzny, D., Marie, E., Metzker, M., Lee, S., Abramson, S., Adams, C., Alder, J.,			
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,			
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,			
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,			
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,			
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,			
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,			
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,			
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,			
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,			
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,			
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,			
Egan, A., Escotto, W., Eugene, C., Evans, C.A., Falls, T., Fan, G.,			
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,			
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,			
Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,			
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,			
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,			
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,			
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,			
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,			
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,			
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,			
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,			
Lorenshuwa, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J.,			
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,			
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,			
Mahoney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,			
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,			
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,			
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,			
Nwankwelu, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,			
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,			
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,			
Puro, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,			
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,			
Rives, C., Rodkey, T., Rojas, R., Rose, R., Ruiz, S. J.,			
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,			

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falle, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gearegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hayes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Louis, S., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., McWhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Sma's, D., Sneed, A., Soderstrom, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

Direct Submission
 Unpublished
 2 (bases 1 to 288118)
 Worley, K. C.

Direct Submission
 Submitted (27-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Oct 11, 2002 this sequence version replaced gi:21998925.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GRMC
 Center clone name: CH230-293111
 ----- Summary Statistics
 Assembly program: Phrap; version 0.930329
 Consensus quality: 214589 bases at least Q40
 Consensus quality: 222817 bases at least Q30
 Consensus quality: 227320 bases at least Q20
 Estimated insert size: 252471; sum-of-contigs estimation
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 5778: contig of 5778 bp in length
 5779 5878: gap of unknown length
 5879 16373: contig of 10495 bp in length
 16374 16473: gap of unknown length
 16474 90420: contig of 73947 bp in length
 90421 90520: gap of unknown length
 90521 236782: contig of 146262 bp in length
 236783 236882: gap of unknown length
 236883 250122: contig of 13240 bp in length
 250123 250222: gap of unknown length
 250223 250333: contig of 6811 bp in length
 250334 257133: gap of unknown length
 257134 259071: contig of 1938 bp in length
 259072 259171: gap of unknown length
 259172 261008: contig of 1837 bp in length
 261009 264448: contig of 3340 bp in length
 264449 264548: gap of unknown length
 264549 267548: contig of 3000 bp in length
 267549 267649: gap of unknown length
 267649 271483: contig of 3834 bp in length
 271483 271583: contig of 4040 bp in length
 271583 275623: contig of 4040 bp in length
 275623 275723: gap of unknown length
 275723 286108: contig of 10385 bp in length
 286108 286208: contig of unknown length
 286208 288118: contig of 1911 bp in length.

FEATURES
 source
 1. .288118
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-293111"
 misc_feature
 1. .1287
 /notes="wgs_end_extension
 clone_end:T7"
 misc_feature
 1338. .5778
 /notes="wgs_end_extension
 clone_end:T7"
 misc_feature
 3086. .3988
 /notes="clone_boundary
 clone_end:T7
 misc_feature
 site:NbOI
 end_sequence:RXAIQ54TJB"
 misc_feature
 5879. .7959
 /notes="wgs_contig"
 misc_feature
 12280. .13622

	D	b	163292	AAC	TCCAGAGAGCCCCAGATTGGTGGCTTTTGACTGGCGCCTAGTATTCACGTGGCACG	TG	163293
	Q	y	948	G	TCTCTAGAGGGAGAGATACGATGCAATGCCCGTCGATGTCATCAGGTCTCCAACA	1007	
	D	b	163232	G	TCCCTTGGCGGAGCGGAAGTTGATGCGGAACTCTATTGCATTATCAGGTCTCCA	163173	
	Q	y	1008	A	TGCTGTGTGGCTGTTTGTCTGAGTAAGAATAATTTTGAATATCTGGGGTCTTATG	1067	
	D	b	163172	A	TGGCTGGGGACTGTTTGTGTGAGTAAGAGATATTTGAATACCTGGGGTCTTATG	163113	
	Q	y	1068	A	CAGGAATGGAAGTTTGGGAGAGAAAACCTCGAATTTTCTTTAGGATCTGGCAGT	1127	
	D	b	163112	A	CCGGAATGGAAGTCTGGGAGSAGAAAACCTCGAATTTCTCTTAGGATCTGGCAGT	163053	
	Q	y	1128	G	TGGGGTCTTGAAAAACACACCATGTTCCCATGTGTGGCCATGTTTCCCACAGCA	1187	
	D	b	163052	G	GT--GACTCAGGAAAACACACCTTATCTCCACGTGGGCCATTTCTTCCCTA	162995	
	Q	y	1188	C	CTACTCCCGCAACAAGGCTCTGGCCAACAGTGTGTGTGAGCTGAAGTATGGA	1247	
	D	b	162994	C	CTATTACGACGACCAAGGCTTAGCCAAACAGTGTCCAGCTGCAGAAAGTGTG	162935	
	Q	y	1248	G	AATTTAAAGAGCTTACTACCATCGCAACCCCGTGGCCCTTGGAACTTTTGGG	1307	
	D	b	162934	G	AATTTAAAGAACTTACTACACCCCGCAAAACCCC-AGSCCCGCTTGGAA	162876	
	Q	y	1308	G	TGACAGAGAGGAAGCAGCTCCGGGACAAGTCTCCAGTGTAAAGACTTCAA	1367	
	D	b	162875	C	TGACAGAGAGAAAGAGCTTCAGCCNAGCTCCAGTGTGAAGACTTNA	162816	
	Q	y	1368	G	AGACTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCTTCTTCGGATG	1427	
	D	b	162815	G	CTACAATGTACCCAGAACTGCATGTGTCAGAGACAGGCCCTGTCTTCCG	162756	
	Q	y	1428	C	AGAACAAAGGACTAAACAGACTACTGTTTTGAGCTATAACCTCCGATGA	1487	
	D	b	162755	G	AGAACAGAGACGACGGGGTACTGCTGAGCTATAATCTCCAGTGA	162696	
	Q	y	1488	G	TGGGACACCAAGTCAATCTCTACTCTGTATCGGATGGCCAGAACTCA	1547	
	D	b	162695	G	AGGCCACCAAGTCTTTTGTACTCTGCCACAGATGGGTCA	162636	
	Q	y	1548	T	ACAGTCCCAAGAAAGAAATACGCTATAACCAACCAAGCTGAGGGTGA	1607	
	D	b	162635	T	GTATATCCCGGCAAGAAATACGGTATAACACCCGACCCAGCTCTG	162576	
	Q	y	1608	G	AACGAGGAATGGATACCTTATCATGATCTTGGAGAGAACTGCCCCAG	1667	
	D	b	162575	G	AGGAAGGAGAGGAT--CCTTATTTCCGGATCTCTGGAAGATACGG	162518	
	Q	y	1668	A	AGTTTCATCTTGACGAGGAGTGA---TCTTTTATTTACGAAACAG	1723	
	D	b	162517	G	AGTTTCATCTTACAGGAGGACGACGACGCTTAGTCTAACGACAG	162458	
	Q	y	1724	C	AGGCTGCGAGGAGGAGTCSAGTGCAGTTTCGTTCCACTCTTACG	1783	
	D	b	162457	G	AGGCCACAGAGAGGAGGATCAACTCTGCCTTTGCTCCGAACTT	162398	
	Q	y	1784	C	TGCGATCATCAGAAATGTGTTCTTTCAAAGAGCGCATGTTATGA	1843	
	D	b	162397	C	TACAGACGCCAGAGGTGGTTCTGCCAGGAGCTCATGTCATAGT	162338	
	Q	y	1844	A	GCCCATCTGAGGAGACTGTGGAGCCAGGACTCTGCCCAACAA	1903	
	D	b	162337	G	AGCAGCCCAACCCGCGCGGGGCTGTGTCC-----CAGTGAG	162305	
	Q	y	1904	G	ACCAGAAACCCCAAAAAATAGGCTGCATTTGTTGAAGAGGCA	1963	
	D	b	162304	--	-----CAGTGAGGAGGACAAAGTACGTGGTTGTCTCT	162272	
	Q	y	1964	G	TGAAAGTGTGTGGATTTAGTAAAAATAGTAAATAGCTTTGT	2023	
	D	b	162271	G	TGAA----TGTGTGGACTCCACAACAAATGTGAATATATTTG	162215	

QY	1488	GTGGGACACAGGTCATCTCTGTATCGGATGGCCAGAAATCAGTTTTTCGAG	1547
Db	162695	GAAGGCCACCAAGTCCCTTTGTACTCTCTGCCACGAGATGGGTCAAGAACAGTTTTTCGAG	162636
QY	1548	TACACGTCCCAAGAAAGAAATACGCTATAACACCCACGAGCTGAGGGCTGCATTGCTGTG	1607
Db	162635	TGTHATCCCGCGAAGAAATACGGTATAACCCGACGCCCAAGGTCGTATAGCAGTG	162576
QY	1608	GAAGCAGGAATGGATACCCTTTATCATGCATCTCTGCGAAGAAATCTGCCCCACAGAGAATCAG	1667
Db	162575	GAGGAAGGAAGGAT--CCTTATTCCGGATCTCTGGAAAGATACGGTTCCAGAGAAATCAA	162518
QY	1668	AAGTTTCATCTTGACGAGGATGGA-----TCTTTATTTTCACGAACAGTCCAAAGAAATGTGT	1723
Db	162517	GAGTTTCATCTCTACAGGAGGACGACAGCAGTGTACTTCAAGCAGACCGGGAATGTGT	162458
QY	1724	CCAGGCTGCGAGGAAGAGAGTCGAGTGACAGTTTCGTTTCCACTCTTTACGAGACTGCACCAA	1783
Db	162457	GGAGGCCACAGAGAGAGATTCAACTCTGCCTTTTGCTCCGAACITTCGAGACTGTAGCNA	162398
QY	1784	CTCGGATCATCAGAAATGGTTCTTTCAAAGAGCGCATGTTATGAAGCCTCGTGATCAAGG	1843
Db	162397	CTCAGACAGCCAGAGTGGTTCTGCCAGGAGCTCATGTCTAGTGCAGGGCCTTGTCAAG	162338
QY	1844	AGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAGCTAAGCAGT	1903
Db	162337	GAAGCAGGCCACCGCGCCGGGGCTGTGTCC-----	162305
QY	1904	GACCAGAAACCCACAAAAATAGGCTGCATTGCTTTGAAGAGGCAATCATTTTGCCATTT	1963
Db	162304	-----CAGTGAGAGGACAAAGTACGTGCTGTCTATCT	162272
QY	1964	GTGAAAGTGTGTTGGATTTAGTAAAAATGTGAATAAGCTTTGTACTTTATTTTGAGAACT	2023
Db	162271	GTGAAA---TGTTGTTGGACTCCACAACAATGTGAATATATTGTACAGATTGGAAACT	162215

```
QY 2024 TTTTAAATGTTCCAAATACCTATTTTCAAGGGTAATCGTAAGATGTTAAACCTTGGT 2083
D 162214 TTTAAATATGCTTCCAGTATCCCGTTTCTAAGGCGGCTAGATGCTGCCCTTGGTA 162155
QY 2084 ATTTAGAAATTAACACCTTATATATTTTCTATCAARAWAWATTTTACAGTCGTGC 2143
D 162154 TTTGGGAAGCTTAATGCACGATGTCAT-----TTTACCCATGTCG 162113
QY 2144 CTTTTACTCTCATTTAGCAAAAAGATTAAGATTTTATTTTGGTATTTTACAGAAATTTCCA 2203
D 162112 CTTTTATTTCTCATTTAGCAAAAAGGGAATATTTTACTTTTGTATTTACGAAGCTTCCCG 162053
QY 2204 GGTACGAAGATATCTGATGGGTGGAATCAGGTTCAAGCAACGCTACTTTTGCATTAAGT 2263
D 162052 GGTGTTTATATATCTTGA-----CACAACTGGTAATACCTCAAAATGGGACTG 162004
QY 2264 ATATACCTCA--GCTGGGGGTTAAAGTTTTCACAGTATAGAGAGACTGTCACCTAGAA 2321
D 162003 GTATACCTCAATGTCATCAAGTTAGAGTTCTCTCAGGGGATGAGATGGACACTAGAA 161944
QY 2322 CATTTGATGATTTATTCAG-----GTCATTTAGATCTTCTAGATGTTATTTTAAAGA 2375
D 161943 TACTGTCTTAATTTCTGTTGTTGTTGAGTGGGTGGGACAGTCTAGATTTTATCAAGAAGA 161884
QY 2376 ATGCTTTTGGTTATGTTGTTGCTACACAGTTAACTC---CATAATGTTTCAATGTCAGC 2432
D 161883 ATGCTTTTGGTTATGTTGTTGCTATGCCACAGTCAAACTCCATCAATACGCTCAGACTGAC 161824
QY 2433 CAAAGAGACTTAACAAAGCTGAATCTCAGAG--AACAAATTTGCTTTTAAAGCTGAGTC 2491
D 161823 CCAACAGGCTCACCATCTCTGAAAGCTCAAGAAAGATATGTTTGGCCAGGCTCTAC 161764
QY 2492 AACTT---GAGAGCGAATCTTCAATGCC-----GCATGATGTTGTCGCTG 2536
D 161763 AAATCTGCGAGTACATTTCCAGCAATGTTATGGGAGGATGTTAGTGAATTTAGGAGG 161704
QY 2537 GTTCTACCATCATGACTTTAAACATGTTTATATCATTTTAAATTTT-ATGATACGTA 2595
D 161703 ACTTAAGACTGTGACTGGAAGAAAGCATTCATATCCTCTAGTCTTAAATCAAGACA 161644
QY 2596 GTGTGAGGAGAAATGAATGTTCTATATGAATTCCTTTTCAAGTTTGTTCATTAATA 2655
D 161643 GTTCCAGGAGAAACGTAATGTTCTACATGAATCACTCAATTTAAAGTTTTCCTAAATA 161584
QY 2656 ACAGTTATTAATTA 2671
D 161583 CTATTTGATTTGTTAA 161568

RESULT 14
AC099237_3
WPCOMMENT
Sequence split into 6 fragments LOCUS AC099237 Accession AC099237
Fragment Name Begin End
AC099237_0 1 110000
AC099237_1 100001 210000
AC099237_2 200001 310000
AC099237_3 300001 410000
AC099237_4 400001 510000
AC099237_5 500001 545260
Continuation (4 of 6) of AC099237 from base 300001 (AC099237 Rattus norvegicus clone CH2

Query Match 31.1%; Score 887.4; DB 2; Length 110000;
Best Local Similarity 69.6%; Pred. No. 8.1e-135;
Matches 1556; Conservative 0; Mismatches 551; Indels 128; Gaps 21;

QY 468 TATGATTTATGATTAATTTCCAGGACATCTGTTATCATAGCATTTTAAATGAAGCCTGG 527
D 43593 TATGATTTATGATTAATTTCCAGGACATCTGTTATCATAGCATTTTAAATGAAGCCTGG 43552
QY 528 TCAACTCTCTCTCGGACAGATTTTACAGTGTCTCTTGAGACATCCCGGATATCTCTGTAGAA 587
```

```
D 43653 TCACACTCTCTTCAGACAGTTTACAGTGTTCGAGACTTCCCTCTGATATCTCTGCTGGAG 43712
QY 588 GAAGTATCTCTCTAGATGACTACAGTGTAGAGACACCTGAAGAGCGCTTTGGCCAAT 647
D 43713 GAGGTCAATCTGTGTAGATGACTACAGTGTAGAGATCACTGAAGAGCGCTTTGGCCA- 43771
QY 648 GAGCTTTTGGGACTGCCCCAAGGTGCGCTGATCCGCCCAACAAGAGAGAGGCGCTGGTG 707
D 43772 GAGCTGTACACAGCTGCCCTAGGTGTACCTGTATCCCTGGTAGCA--AGAGAGGCGCTAGTG 43829
QY 708 CGAGCCCGGCTGTCTGGGGGGTCTCGGCGAGGGGGGATGTTCTGACCTTCTCTGGACTGT 767
D 43830 CAAG-CCGGCTACTTGGGAGGCTCTGGGGCCAGGGGGAAGTGTGAGGTTTCTGGACTGT 43888
QY 768 CACTGTGAGTGCACCAAGGCTGCTGGAGCGCTGTCGAGAGGATCCATGAAGAGGAG 827
D 43889 CACTGTGACTGTCTAGGGGGTGGCTGGAGT---CTGCTGAGAGGATCCACGAGAGGAG 43946
QY 828 TCGGCAGTGTGTGCCCGGTGATTGATGTGATCGACTGGAACACCTTCGAATACCTGGGG 887
D 43947 TCGGCGTGTGTGCCCGGTGATTGATGTGATCGACTGGAACACCTTCGATGACTTGGGC 44006
QY 888 AACTCGGGAGGCGCCAGATCGCGGTTCGATCGAGGCTGTGTGTCAGTGCAGCAGACA 947
D 44007 AACTCC--GGAAGCCCGAGATTTGAGGCTTTGACTGGCGCTAGTATTCAGTGCAGCGT 44065
QY 948 GTTCTCGAGAGGAGAGATACGATGCAATCCCGCTGATGTCATCAGGTCTCCAACA 1007
D 44066 GTCCTTTGGCGGAGCGGAAGTTGATGCGGACTCTTATGACATTCAGGTCTCCAAC 44125
QY 1008 ATGGCTGTGGGCTGTTTGTGTGAGTAAGAAATATTTTGAATATCTGGGCTCTTATGAT 1067
D 44126 ATGGCTGTGGGCTGTTTGTGTGAGTAAGAAATATTTTGAATATCTGGGCTCTTATGAT 44185
QY 1068 ACAGGAATGAAATTTGGGAGGAGAAACCTCGAATTTTCTTTTAGGATCTGGCAGTGT 1127
D 44186 ACCGAATGAAATCTAGGAGGAGAAACCTTGAATTTCTTTTAGGATCTGGCAGTGT 44245
QY 1128 GGTGGGTTCTGGAACACACACACCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCT 1187
D 44246 GGTGACTCTTGGAAACACACACCTTACTCCACGTTGGGCCACGCTCTTCCCTAGCAAGCT 44305
QY 1188 CCCTACTCTCCGCAACAGGCTCTGGGCCAACAGTGTTCGTGCGAGCTGAAGTATGATGGAT 1247
D 44306 CCCTATTCAGCAGCAAGGCTCTAGCCAACAGTGTCCGASCTGCAGAAGTGTGGCGGAT 44365
QY 1248 GAATTTAAAGAGCTCTACTACATCGCAACCCCGCTGCCGCTTGGAACTTTTGGGGAT 1307
D 44366 GAACTTTAAAGAGCTCTACCACCCCGCAAAATCCCC--AGCCCGCTGGAAACCTCTTGGGGAG 44424
QY 1308 GTGACAGAGAGGAGAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGTCTTCTG 1367
D 44425 CTGACAGAGAGAAAGAGCTTCAGCCCAAGCTCAAGTGTGAAGATTTAAAGCGGTTCTCTG 44484
QY 1368 GAGACTGTGTATCCAGAACTGTCATGCTGTCGAGGACAGGCTGCTTCTTCTCGGATGCTC 1427
D 44485 GCTACACTGTACCCAGAACTGTCATGTCAGAGAGAGAGGCTGCTTCTTCTCGGAGCTT 44544
QY 1428 CAGAACAAAGAGCTTAACAGACTACTGCTTTTGTATATAACCTCCCGATGAACACAGATT 1487
D 44545 GAGAACAGAGACTTACGGGGTACTGCTGGAATATATAATCTCTCCAGTGAACAAATGTG 44604
QY 1488 GTGGGACACAGGCTCATTTCTGACTCTGTCATGGGATGGCCAGATTCAGTTTTCGAG 1547
D 44605 GAAGGCCACAGGCTCTTTTGTACCTCTGCCAGGGATGGGTCAAGAACCAAGTTTTCGAG 44664
QY 1548 TACAGCTCCAGAGAAATAACGCTATAACCCACAGGCTTGAAGGCTGCTGCTGCTGTG 1607
D 44665 TGTACATCCAGCAGAAATAACGCTATAACA--TCGCCAGCCCGCAGGCTCTGATAGCAGTG 44723
QY 1608 GAAGCAGGAATGGAATCCCTTATCATGCTCTCTGCGAAGAACTGCCCGCAGAAATCAG 1667
D 44724 GAGGAAGGAAGGAT--CCTTATCCCGGCTCTCTGGAAGATATACCGTTCCAGAGAAATCAA 44781
```


Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GTSQ
 Center clone name: CH230-227M12
 ----- Summary Statistics

Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 142428 bases at least Q40
 Consensus quality: 149083 bases at least Q30
 Consensus quality: 153531 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 33 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1215: contig of 1215 bp in length
* 1216: gap of unknown length
* 1316: contig of 1512 bp in length
* 2827: gap of unknown length
* 2927: gap of unknown length
* 4164: contig of 1237 bp in length
* 4264: gap of unknown length
* 4265: contig of 1563 bp in length
* 5827: gap of unknown length
* 5828: contig of 2730 bp in length
* 8657: gap of unknown length
* 8658: contig of 1208 bp in length
* 8758: gap of unknown length
* 9966: gap of unknown length
* 10065: contig of 2717 bp in length
* 12782: gap of unknown length
* 12783: gap of unknown length
* 12882: gap of unknown length
* 14584: gap of unknown length
* 14585: contig of 2414 bp in length
* 14684: gap of unknown length
* 17098: gap of unknown length
* 17198: gap of unknown length
* 18477: contig of 1279 bp in length
* 18478: gap of unknown length
* 21168: contig of 2591 bp in length
* 21268: gap of unknown length
* 23708: contig of 2440 bp in length
* 23808: gap of unknown length
* 26530: contig of 2722 bp in length
* 26531: gap of unknown length
* 26630: gap of unknown length
* 28801: contig of 2171 bp in length
* 28901: gap of unknown length
* 33091: contig of 4190 bp in length
* 33191: gap of unknown length
* 36968: contig of 3777 bp in length
* 37068: gap of unknown length
* 39304: contig of 2236 bp in length
* 39404: gap of unknown length
* 43433: contig of 4029 bp in length
* 43534: gap of unknown length
* 46236: contig of 2703 bp in length
* 46336: gap of unknown length
* 50699: contig of 4363 bp in length
* 50799: gap of unknown length
* 53947: contig of 3148 bp in length
* 54047: gap of unknown length
* 58293: contig of 4246 bp in length
* 58393: gap of unknown length
* 63160: contig of 4767 bp in length
* 63260: gap of unknown length
* 68334: contig of 5074 bp in length
* 68435: gap of unknown length
* 73462: contig of 5028 bp in length

```

```

* 73463 73562: gap of unknown length
* 73563 79617: contig of 6055 bp in length
* 79618 86568: contig of unknown length
* 86569 86568: contig of 6851 bp in length
* 86569 86568: gap of unknown length
* 95652 95652: contig of 8984 bp in length
* 95653 95752: gap of unknown length
* 95753 105300: contig of 9548 bp in length
* 105301 105400: gap of unknown length
* 105401 112938: contig of 7538 bp in length
* 112939 113038: gap of unknown length
* 113039 122447: contig of 9409 bp in length
* 122448 122547: gap of unknown length
* 122548 137894: contig of 15347 bp in length
* 137895 137894: gap of unknown length
* 137995 166547: contig of 28553 bp in length.
FEATURES
    Location/Qualifiers
        source
            1..166547
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-227M12"
BASE COUNT 45529 a 35358 c 35370 g 45112 t 5178 others
ORIGIN
Query Match      31.1%; Score 887.4; DB 2; Length 166547;
Best Local Similarity 69.6%; Pred. No. 8.5e-135;
Matches 1556; Conservative 0; Mismatches 551; Indels 128; Gaps 21;
QY 468 TATGATTATGATAAATTTGGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 527
Db 84617 TATGATTATGATAAATTTGGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 84558
QY 528 TCAACTCTCTTCGGACAGTTTACAGTCTTTCAGACATCCCGGATATCTGCTAGAA 587
Db 84557 TCACACACTCTTCAGACAGTTTACAGTCTTTCAGACATTCCTGATATCTGCTGAG 84498
QY 588 GAAGTGATCTCTGTAGATGACTACAGTATAGAGAGACCTGAAGAGGCGCTTGGCCAAAT 647
Db 84497 GAGGTATCTCTGTAGATGACTACAGTATAGAGAGACCTGAAGAGGCGCTTGGCCAA - 84439
QY 648 GAGCTTTGGGAGATGCCCCAGGTGCGCTGATCCGCCCAACAAGAGAGAGGCGCTGGTG 707
Db 84438 GAGCTGTACAGCTGCGCTTACCTAGGTGTACCTGATCCCTGGTAGCA--AGAGAGGCGCTAGTG 84381
QY 708 CGAGCCCGGCTGCTGGGGGGCTCTGGCGAGGGGAGTCTTCTGACCTTCTGCGACTGT 767
Db 84380 CAAG-CGCGCTACTGGGAGCCTCTGGGGCCAGGGGCGAAGTGTGAGGTTTCTGAGACTGT 84322
QY 768 CACTGTGAGTGCCACGAAGGGTGGCTGGAGCCGCTGCTGCAGAGGATCCATGAAGAGGAG 827
Db 84321 CACTGTGAGTGTCTATGAGGGTGGCTGGAGT--CTGCTGCAGAGGATCCACGAGAGGAG 84264
QY 828 TCGGCACTGCTGTGCCCGGTGATGATGATCGATCGATCGGAAACCTTCGAATACCTGGGG 887
Db 84263 TCGGCGGTGTGTGCCCGGTATCGATGTCTGACTGGAACACTTCTGAGTACCTGGGC 84204
QY 888 AACTCCGGGAGGCCCGAGATCGCGGGTTTCGATGAGGGCTGGTGTTCAGTGGCACACA 947
Db 84203 AACTCC--GGAAGGCCCGAGATTCGAGGCTTTGACTGGCGCCTAGTATTTCAGTGGCAGCTG 84145
QY 948 GTTCTCTGAGAGGAGGAGGATACGGATGCAATCCCGCTCGATGTCACTCAGGTCTCCAAACA 1007
Db 84144 GTCTCTTGGCGGAGGCGGAGTGTGATGCGGACTCTTATGACATTATCAGGTCTCCAACT 84085
QY 1008 ATGCTGTGCGGCTGTTTGTGTGAGTAAAGAAATATTTTGAATATCTGGGGTCTTATGAT 1067
Db 84084 ATGCTGTGCGGCTGTTTGTGTGAGTAAAGAAATATTTTGAATACCTGGGGTCTTATGAT 84025
QY 1068 ACAGGATGGAAGTTTGGGAGGAGGAGAAACCTCGAATTTTCCCTTAGGATCTGCACTGT 1127
Db 84024 ACCGGAATGGAAGTCTAGGAGGAGGAGAAACCTTGAATTTCTCCTTAGGATCTGCAATGT 83965

```

```
QY 1128 GGTGGGTTCTGGAAACACACCCATGTTCCCATGTTCCCATGTTTCCCAAGCAGCT 1187
Db 83964 GGTGACTCTCTGNAACACACCCCTTACTCCCATGTTGGCCACAGCTCTTCCCTTAAGCAAGCT 83905
QY 1188 CCTACTCCGCAACAAAGGCTCTGGCAACAGTGTTCGTGCAGCTGAAGTATGGATGGAT 1247
Db 83904 CCTATTACGCGACGAGGCTCTAGCCCAACAGTGTCCGAGCTGCAGAAGTGTGGCGGAT 83845
QY 1248 GAATTTAAAGAGCTCTACTACCACTCGCAACCCCGGTCCCGCTTGGAAACCTTTTGGGGAT 1307
Db 83844 GAATTTAAAGAGCTCTACTACCAACCCCGGAAATCCCC - AGCCGCGCTGGAAACCTCTCGGGAG 83786
QY 1308 GTACAGAGAGGAGCAGCTCCGGGCAACGCTCCAGTGTAAAGACTTTCRAAGTGTCTTCTG 1367
Db 83785 CTGACAGAGAGAGAGAGCTTTCAGCCCAAGCTTCAGTGTGAAGATTTAAAGCGGTCTCTG 83726
QY 1368 GAGACTGTGTATCCAGAACTGTCATGTGCTGAGACAGGCTGCTCTTTCGGGATGCTC 1427
Db 83725 GCTACACTGTATCCAGAACTGTCATGTGCTGAGACAGGCTGCTTCTTCGGGAGCTT 83666
QY 1428 CAGAACAAAGACTTAACAGACTACTGCTTTGACTATTAACCTCCCGATGAAACCAAGATT 1487
Db 83665 GAGAACAGAGAGACTTACGGGGGTACTGCTCGGACTATAATCTCCAGTGAACCAATGTG 83606
QY 1488 GTGGGACACCAAGTCTTCTGTACCTCTGTGATGGGATGGCCAGATCAATTTTCGAG 1547
Db 83605 GAAGGCCACCAAGTCTTCTGTACCTCTGCTGCGGATGGGTGAGAACCAAGTCTTTCGAG 83546
QY 1548 TACACGCTCCAGAGAAATACGCTATAACCCACAGGCTCAGGGCTCATTGTGCTG 1607
Db 83545 TGTACATCTCCAGAGAAATACGCTATAAC - TCGCAGGCCAGGCTCTGCATAGCAGTG 83487
QY 1608 GAAGCGGAATGATACCTTTATCATGCTATCTCTGCGAAGAACTGCCCCAGAGAAATCAG 1667
Db 83486 GAGGAAGGGAAGGAT - CCTTATCCCGGGTCTCTGGAAGATACCGTTCAGAGAAATCAA 83429
QY 1668 AGTTCATCTTGAGGAGGATGATCTTTATTTTCAAGACAGTCCAGAAATGTGCCAG 1727
Db 83428 GAGTTCATCTTACAGAGGACCGGACCTTAGTCTTCAAGCAGAGCAGGAAATGTGTGGAG 83369
QY 1728 GCTGCGAGGAGGAGTGCAGTGCAGTTCGTTCCACTTTACGAGACTGCACCAACTCG 1787
Db 83368 GCCACAGAGGAGGAGTTCAACTCTGCTTGTCTCCGAACTTGGAGACTGTAGCAACTCA 83309
QY 1788 G - ATCATCAGAAATGGTCTTCAAGAGCGCATGTATGAAGCCTCGTGTATCAAGGAG 1845
Db 83308 GACACAGCCAGAGGTGGTCTTCAAGGAGTGCATGTATAGTGCACGGCTTGTGAGG - 83251
QY 1846 CCATCGAGGAGACTGTGAGGACGAGCTCTGCCCAACAAAGACTTAGCTAAGCAGTGA 1905
Db 83250 - - - - - GAAGGAGCCACCGCGCGCGGGGCTG - - - - - 83224
QY 1906 CCAGAACCCACCAAAACTAGGCTGATGCTTTTGAAGGCGCAATCATTTTGCATTTGT 1965
Db 83223 - - - - - TGTCCCACTGAGAGGAGCAAGTACGTGGTGTGTCATCTGT 83183
QY 1966 GAAAGTGTGTGGATTTAGTAAAAATGTAAGCTTTGTACTTATTTTGAGAACTTT 2025
Db 83182 GRA - - - TGTGTGGACTCCACACAAATGTAATTTATTTGTACAGATTTGGAACCTTT 83126
QY 2026 TTAATGTTCAAAATACCTTATTTTCAAGGGTAAATCGTAAGATGTAAACCTTGGTAT 2085
Db 83125 AAAAAATGTCTTCCAGTATCCGTTTCTAAGGGCGCTCTAAGATGCTGCTTTTGGTAT 83066
QY 2086 TTAGAAAAATTAACCTTATAATTTTCTATCAARAWAWATTTTACAGTCTGCTCT 2145
Db 83065 TGGGAAGCTTATGACAGATGTCAT - - - - - TTACCATGGTGCCT 83024
QY 2146 TTACTCTCTATTAGCAAAAAGATAAAGATTTTATTTTGGTATTTTACAAGAAATTCACAGG 2205
Db 83023 TTTATTTCTATTAGCAAGAGGGAATATTTTACTTTGTACTTACGAGCTTCCGGG 82964
QY 2206 TACGAAGATATCTGCAATGGGTGGAATACAGGTTCAAGCAACGATCTTTGCAATTAACATGAT 2265
```

```
Db 82963 TGTGTAGTATACTTGA - - - - - CACAACTGGTAATACCTCAAAATGTGGACTGGT 82915
QY 2266 AATACCTCA - - GCTGGCGGGTTAAAGTTTCCAGTATATAGAGAGACTGTCACTAGGAACA 2323
Db 82914 AATACCTCAATGTCATCAAGTTAGAGTCTCTCAGGGGATGAGATGACACTAGAATA 82855
QY 2324 TTGTATTGATTATTTCAG - - - - - GTCATTTGAGATCTTCTAGATGTATTTTAAAGAAAT 2377
Db 82854 CTGTCTTAATTTGTTTGTGTGTGAGTGGGACAGTCTAGATTTTATACAAGAAAGAT 82795
QY 2378 GCTTTTGGTTATGCTGTGCTACCAAGTAAACACTC - - - CATAAATGTTTCATGTCAGCCA 2434
Db 82794 GCTTTTGTCTTATGCTGTGATGCCAGATCAAAACCTCCATCATACGCTCAGACTGACCC 82735
QY 2435 AAGAGGACTAAACCAAGCTGAAATCTCAGAG - AACAAATTTGCTTTTACTAAAGCTGAGTCAA 2493
Db 82734 AAGAGGCTCACCATCTCTGAAAGCTCAAGAAAGAAATATGGTTTGCCAGGTCCTACAA 82675
QY 2494 CTT - - - GAGAGCGAACTTCTAAACAATGCCGCACTGTAGTGTGCTGCT - - - - - 2538
Db 82674 ACTTGGCAGTGTACATTTCCAGCAATGTCTAGTGGGAGGATGTCTAGTGACTTACGAGGA 82615
QY 2539 -TCTACCACTATGACTTTAAACATGTTTATATCATTTTAAATTTT - ATGATACGGTAG 2596
Db 82614 CTTAACGACTGTGACTGAAAGAAAGCAATTCAGATCATTTCTAGTTTAAATCAAAGGACAG 82555
QY 2597 TGTCCAGGAGAAATGTAATGTTTCTATATGAAATTCCTTTTCAAGTTTGTTCATTAATAA 2656
Db 82554 TTCCAGGAGAAACGTAATGTTTCTACATGAACCTCGACTTAAAGTTTTTCTTAATAA 82495
QY 2657 CAGTTATTAATTAA 2671
Db 82494 TATTGATTGTTAA 82480
```

Search completed: December 7, 2003, 18:37:45
Job time : 10426.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 11:24:22 ; Search time 736.064 Seconds
(without alignments)
10452.079 Million cell updates/sec

Title: US-10-074-527-1
Perfect score: 2850
Sequence: 1 cgggctcggtaccactataa.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Genes19Jun03.*

```
1: /SIDSL1/gcgdata/geneseq/geneq-emb1/NA1980.DAT.*
2: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1981.DAT.*
3: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1982.DAT.*
4: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1983.DAT.*
5: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1984.DAT.*
6: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1985.DAT.*
7: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1986.DAT.*
8: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1987.DAT.*
9: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1988.DAT.*
10: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1989.DAT.*
11: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1990.DAT.*
12: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1991.DAT.*
13: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1992.DAT.*
14: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1993.DAT.*
15: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1994.DAT.*
16: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1995.DAT.*
17: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1996.DAT.*
18: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1997.DAT.*
19: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1998.DAT.*
20: /SIDSL1/gcgdata/geneq/geneq-emb1/NA1999.DAT.*
21: /SIDSL1/gcgdata/geneq/geneq-emb1/NA2000.DAT.*
22: /SIDSL1/gcgdata/geneq/geneq-emb1/NA2001A.DAT.*
23: /SIDSL1/gcgdata/geneq/geneq-emb1/NA2001B.DAT.*
24: /SIDSL1/gcgdata/geneq/geneq-emb1/NA2002.DAT.*
25: /SIDSL1/gcgdata/geneq/geneq-emb1/NA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2848	99.9	2850	ABSS2469	Human cDNA encodin
2	2738.2	96.1	2745	AA040568	Human drug metabol
C 3	2081.8	73.0	2290	AA026522	Human breast cancer
C 4	2081.8	73.0	2290	AA026538	Human breast cancer
5	1208.4	42.4	1259	AA098209	Human colon cancer
C 6	1115.2	39.1	1352	ABV23465	Human prostate exp
C 7	1115.2	39.1	1352	ABV29323	Human prostate exp
8	1012.2	35.5	1517	AAH33532	Human colon cancer

same as in gene emb1

9	1012.2	35.5	1517	24	ABL90290	Human polynucleoti
10	1011.8	35.5	1517	21	AAA78414	DNA encoding novel
11	967.4	33.9	1069	23	AA568954	Human breast cancer
C 12	756.2	26.5	883	22	AA026610	Human breast cancer
13	552.8	19.4	592	21	AA026624	Human colon cancer
14	527	18.5	554	22	AAH82814	Human ovarian tumor
15	526.4	18.5	544	22	AA524233	Human ovarian PCR-
16	497	17.4	652	22	AAH82556	Human ovarian tumor
17	496.4	17.4	632	22	AA523991	Human ovarian PCR-
18	496.4	17.4	632	22	AA525242	Human ovarian tumor
19	496.2	17.4	1737	25	ABX14962	Human N-acetylglala
20	450.8	15.8	496	24	ABU81812	Human ovarian cancer
C 21	445.4	15.6	470	22	AA016682	Human breast cancer
C 22	443.4	15.6	473	24	ABU81811	Human ovarian cancer
C 23	425.8	14.9	473	24	ABU80705	Human ovarian cancer
C 24	425	14.9	437	22	AA025525	Human breast cancer
25	423	14.8	425	23	ABV14650	Human prostate exp
26	423	14.8	471	23	ABV35728	Human prostate exp
27	423	14.8	471	23	ABV45532	Human prostate exp
28	422	14.8	439	23	ABV05481	Human prostate exp
C 29	419.2	14.7	450	22	AA016580	Human breast cancer
C 30	415.4	14.6	417	22	AA025423	Human breast cancer
C 31	409	14.4	421	24	ABU62213	Colon adenocarcino
C 32	409	14.4	421	24	ABU67829	Ovary cancer relat
C 33	406.4	14.3	408	24	ABU87421	Human ovarian cancer
C 34	363	12.7	483	24	ABK53538	Human eosinophil-m
C 35	361	12.7	414	22	AA070779	Human breast cancer
C 36	343.8	12.1	384	22	AA070677	Human breast cancer
37	285.6	10.0	1920	22	AA063512	UDP-GalNac: polype
38	285.6	10.0	2102	24	ABQ99468	Human coding seque
39	285.6	10.0	2575	21	AAA37122	Human PRO1564 (UNQ
40	285.6	10.0	2575	22	AA546106	Human DNA encoding
41	285.6	10.0	2575	22	AA544445	Primer #98 used in
42	285.6	10.0	2575	25	ACA57864	Human PRO1564 CDNA
43	285.6	10.0	2575	25	ABX98334	Human cDNA encodin
44	285.6	10.0	2575	25	ABX98836	Novel human secret
45	285.6	10.0	2575	25	ACA05981	Human secreted/tra

ALIGNMENTS

RESULT 1
ABSS2469
ID ABS52469 standard; cDNA; 2850 BP.
XX ABS52469;
AC ABS52469;
XX ABS52469;
DT 15-NOV-2002 (first entry)
XX Human cDNA encoding glycosyltransferase 33945.

Human; ss; gene; glycosyltransferase; 33945; atherosclerosis;
cardiovascular disorder; ischaemia; atherosclerosis; cancer; tumour;
congenitive heart failure; angina; hypertension; atrial fibrillation;
diabetic neuropathy; angina; hypertension; atrial fibrillation;
valvular disease; cardiomyopathy; haemangioma; pancreatic disorder;
cellular proliferative disorder; differentiative disorder; diabetes;
autoimmune disorder; haematopoietic neoplastic disorder; leukaemia;
Hodgkin's disease; chronic myelogenous leukaemia; inflammatory disease;
arthritis; multiple sclerosis; viral infection; liver disorder;
liver fibrosis; hepatocellular cancer.

OS Homo sapiens.

XX Key Location/Qualifiers
CDS 81..1826
/tag= a
/product= "Glycosyltransferase 33945"
/note= "this CDS is specifically claimed in claim 1"

XX WO200264815-A2.

XX

PD 22-AUG-2002.
XX
XX
XX 14-FEB-2002; 2002MO-US05042.
XX
XX 15-FEB-2001; 2001US-269202P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Olandt PJ, Meyers RE, Galvin KA;
XX WPI; 2002-643491/69.
XX P-PSDB; ABG32510.
XX
XX New 33945 nucleic acid molecule and polypeptide, useful for diagnosing
XX PT and treating disorders involving aberrant or deficient
XX PT glycosyltransferase function or expression such as atherosclerosis or
XX PT endothelial cell disorders -
XX
XX PS Claim 1; Page 97-100; 107pp; English.
XX
XX The invention relates to an isolated 33945 nucleic acid molecule encoding
XX CC a glycosyltransferase polypeptide, its fragment or allelic variant
XX CC and the encoded protein. Also included are host cells, antibodies,
XX CC producing the protein from cell culture, detecting the
XX CC protein/nucleic acid using probes or binding compounds and identifying
XX CC compounds (modulators) which bind the protein. The methods and
XX CC compositions of the present invention are useful for diagnosing and
XX CC treating disorders involving aberrant or deficient glycosyltransferase
XX CC function or expression such as atherosclerosis, cardiovascular
XX CC disorders (e.g. ischaemia, atherosclerosis, congestive heart failure)
XX CC endothelial cell disorders (e.g. tumours, psoriasis, diabetic
XX CC retinopathy) angina, hypertension, atrial fibrillation,
XX CC valvular disease, cardiomyopathy, haemangiomas, cancers, pancreatic
XX CC disorders, cellular proliferative and/or differentiative disorders,
XX CC autoimmune disorders, haematopoietic neoplastic disorders (e.g.
XX CC leukaemia, Hodgkin's disease, chronic myelogenous leukaemia),
XX CC inflammatory diseases (e.g. diabetes, arthritis, multiple sclerosis),
XX CC viral infection (e.g. Hepatitis B, hepatitis C and herpes simplex
XX CC virus) and liver disorders (e.g. liver fibrosis and hepatocellular
XX CC cancer). Many more diseases and disorders are listed in the
XX CC specification. The present sequence encodes the glycosyltransferase
XX CC 33945.
XX
XX Sequence 2850 BP; 743 A; 637 C; 761 G; 704 T; 5 other;

Query Match 99.98; Score 2848; DB 24; Length 2850;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 2850; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGCTCGGTACCACTATAACGGCCGCGCAGTGTGCTGGAATTCGCCCTTGGCAGATCGC 60
DB 1 CCGGCTCGGTACCACTATAACGGCCGCGCAGTGTGCTGGAATTCGCCCTTGGCAGATCGC 60
QY 61 TGGCTGAGTGGCGGCGCATGTGGGGGCGCAGCGCGCGCGCGTGGCGGGGAAAC 120
DB 61 TGGCTGAGTGGCGGCGCATGTGGGGGCGCAGCGCGCGCGCGTGGCGGGGAAAC 120
QY 121 TGGCGCGCGCGCGGAGCGCTGTGTGTCTCTGCGCGCTACTGGCGGCTGGCGGGCTGG 180
DB 121 TGGCGCGCGCGCGGAGCGCTGTGTGTCTCTGCGCGCTACTGGCGGCTGGCGGGCTGG 180
QY 181 GCTCGGTGCTGCGGGCGCAGCTGGGGGCGGGGCGGGGCTGCCAGCCGGGACCCCGCG 240
DB 181 GCTCGGTGCTGCGGGCGCAGCTGGGGGCGGGGCGGGGCTGCCAGCCGGGACCCCGCG 240
QY 241 GCACCCCGCGCCCGGCGGCGAGCCGGTCAATGCCCGCGCGCGCGCGCGCGCGCGCG 300
DB 241 GCACCCCGCGCCCGGCGGCGAGCCGGTCAATGCCCGCGCGCGCGCGCGCGCGCGCG 300
QY 301 CGCTGGGCGCGCGGGGCGAGCGGTGCGGTGCGAGCTGCGAGGGCGAGAGCTGGGCTGC 360
DB 301 CGCTGGGCGCGCGGGGCGAGCGGTGCGGTGCGAGCTGCGAGGGCGAGAGCTGGGCTGC 360

QY 361 AGGAGGAGAGCGTGGCGGTGCACGAGTTAACTACTACTCAGGACCGCATCTCACTGC 420
DB 361 AGGAGGAGAGCGTGGCGGTGCACGAGTTAACTACTACTCAGGACCGCATCTCACTGC 420
QY 421 ACCGCGCGCTGCCGAGCGCTGGAAACCGCTGTGCAAGAGAGAGAAATATGATTATGATA 480
DB 421 ACCGCGCGCTGCCGAGCGCTGGAAACCGCTGTGCAAGAGAGAGAAATATGATTATGATA 480
QY 481 ATTGTGCCAGGACATCTGTTTATCATAGCATTTTATATGAAGCTGGTCAACTCTCTTTC 540
DB 481 ATTGTGCCAGGACATCTGTTTATCATAGCATTTTATATGAAGCTGGTCAACTCTCTTTC 540
QY 541 GGACAGTTTACAGTGTCTTGAGACATCCCGGATATCTCTGTAGAGAGAGTATCTTTC 600
DB 541 GGACAGTTTACAGTGTCTTGAGACATCCCGGATATCTCTGTAGAGAGAGTATCTTTC 600
QY 601 TAGTACATACAGTATAGAGACATCTGAGAGAGCGCTTGGCCATAGCTTTCGGGAC 660
DB 601 TAGTACATACAGTATAGAGACATCTGAGAGAGCGCTTGGCCATAGCTTTCGGGAC 660
QY 661 TGCCCAAGGTGGCGCTGATCCGCGCCAAACAGAGAGAGGCGCTGTGTCGAGCCCGGTGC 720
DB 661 TGCCCAAGGTGGCGCTGATCCGCGCCAAACAGAGAGAGGCGCTGTGTCGAGCCCGGTGC 720
QY 721 TGGGGGCGTCTGCGGCGAGGCGCGATGTTCTGACCTTCTGTGACTGTCACTGTGAGTGC 780
DB 721 TGGGGGCGTCTGCGGCGAGGCGCGATGTTCTGACCTTCTGTGACTGTCACTGTGAGTGC 780
QY 781 ACAGAGGTGGCTGGAGCGCGCTCTGCGAGAGATCCATGAAGAGAGTCCGCGAGTGTGT 840
DB 781 ACAGAGGTGGCTGGAGCGCGCTCTGCGAGAGATCCATGAAGAGAGTCCGCGAGTGTGT 840
QY 841 GCCCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
DB 841 GCCCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
QY 901 CCAGATCGCGCGTTCGACTGGAGGCTGGTGTTCAGCTGGGCGCACACAGTTCCTGAGAGG 960
DB 901 CCAGATCGCGCGTTCGACTGGAGGCTGGTGTTCAGCTGGGCGCACACAGTTCCTGAGAGG 960
QY 961 AGAGATACGAGTGCATCCCGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB 961 AGAGATACGAGTGCATCCCGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1020
QY 1021 TGTTCGTGTGAGTAAAGAAATATTTGAAATATCTGGGTCTTATGATACAGGAATGAAG 1080
DB 1021 TGTTCGTGTGAGTAAAGAAATATTTGAAATATCTGGGTCTTATGATACAGGAATGAAG 1080
QY 1081 TTTGGGGAGGAGAAAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGGGTCTGG 1140
DB 1081 TTTGGGGAGGAGAAAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGGGTCTGG 1140
QY 1141 AAACACACCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCA 1200
DB 1141 AAACACACCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCA 1200
QY 1201 ACAGGCTCTGGCCAAACAGTGTTCGTGACGTGATGATGATGATGATGATGATGATGATGAT 1260
DB 1201 ACAGGCTCTGGCCAAACAGTGTTCGTGACGTGATGATGATGATGATGATGATGATGATGAT 1260
QY 1261 TCTACTACCATCGCAACCCCGCTGCCCGCTTGGAAACCTTTTGGGGATGTGACAGAGGA 1320
DB 1261 TCTACTACCATCGCAACCCCGCTGCCCGCTTGGAAACCTTTTGGGGATGTGACAGAGGA 1320
QY 1321 AGCAGCTCCGGGACAGAGCTCCAGTGTAAAGACTTTCAAGTGGTCTTGGAGACTGTGATC 1380
DB 1321 AGCAGCTCCGGGACAGAGCTCCAGTGTAAAGACTTTCAAGTGGTCTTGGAGACTGTGATC 1380
QY 1381 CAGAACTGCATGTCCTGAGGACAGGCTCGCTTCTTCGGGATGCTCCAGAAACAAAGAC 1440
DB 1381 CAGAACTGCATGTCCTGAGGACAGGCTCGCTTCTTCGGGATGCTCCAGAAACAAAGAC 1440
QY 1441 TAAACAGACTACTGCTTTGACTATTAACCCCTCCGATGAAACCAAGATTGTGGGACACGAG 1500

2521	CACTGTAGTGGCTGGTCTTACCACTATGACTTTTAAAAACATGTTTATATCAATTTTAAAT	2580
2581	TTTTATGATACGGTAGTGTACAGGGAGAAATGTAATGTTCTATATGAAAATTCCTTTTCAA	2640
2581	TTTTATGATACGGTAGTGTACAGGGAGAAATGTAATGTTCTATATGAAAATTCCTTTTCAA	2640
2641	GTTTGTTCATTAATAACAGTTTATTAATTTTAAATCAGCGTTAGAGTTTGTGCTGCGCAAC	2700
2641	GTTTGTTCATTAATAACAGTTTATTAATTTTAAATCAGCGTTAGAGTTTGTGCTGCGCAAC	2700
2701	TGCTGTGAAAATTTCTCTGAGTAAATCTCGATTGTTGTAATGATCCAGACCAACCCCTGAGA	2760
2701	TGCTGTGAAAATTTCTCTGAGTAAATCTCGATTGTTGTAATGATCCAGACCAACCCCTGAGA	2760
2761	TTTTGTCAACCTGATTAAAGTCAATATGAATGAATTAAGAGATGTTGAGAAACAAAAA	2820
2761	TTTTGTCAACCTGATTAAAGTCAATATGAATGAATTAAGAGATGTTGAGAAACAAAAA	2820
2821	AAAAAAAAAAAAAAAAAAAAAAAAAAAAA	2850
2821	AAAAAAAAAAAAAAAAAAAAAAAAAAAAA	2850
<div> <div>RESULT 2</div> <div>AAD40568</div> <div>ID AAD40568 standard; cDNA; 2745 BP.</div> <div>XX</div> <div>AC AAD40568;</div> </div>		
XX	30-OCT-2002 (first entry)	
XX	Human drug metabolising enzyme (DME-4) cDNA.	
XX	Human; drug metabolising enzyme; autoimmune; inflammatory disorder;	
XX	acquired immunodeficiency syndrome; AIDS; atherosclerosis; psoriasis;	
XX	proliferative disorder; arteriosclerosis; cirrhosis; hepatitis; cancer;	
XX	asthma; neurological disorder; Alzheimer's disease; Huntington's disease;	
XX	dementia; Parkinson's disease; developmental disorder; anaemia; adenoma;	
XX	drug screening; endocrine disorder; conjunctivitis; glaucoma; cataract;	
XX	renal tubular acidosis; eye disorder; epilepsy; thrombosis; peptic ulcer;	
XX	anorexia; metabolic disorder; cystic fibrosis; diabetes; liver disorder;	
XX	goitre; gastrointestinal disorder; gene therapy; virucide; anticoagulant;	
XX	anticonvulsant; nootropic; enzyme; DME-4; gene; ss.	
XX	Homo sapiens.	
Key	Location/Qualifiers	
XX	1..1746	
FT	/tag= a	
FT	/product= "Human DME-4"	
FT	1..102	
FT	/tag= b	
FT	103..1743	
FT	/tag= c	
FT	/product= "Mature human DME-4"	
FT	1..117	
FT	/tag= d	
FT	118..1743	
FT	/tag= e	
FT	/product= "Mature human DME-4"	
XX	WO200246426-A2.	
XX	13-JUN-2002.	
XX	04-DEC-2001; 2001WO-US47429.	
XX	08-DEC-2000; 2000US-254308P.	
XX	15-DEC-2000; 2000US-256189P.	
XX	21-DEC-2000; 2000US-257113P.	
XX	19-JAN-2001; 2001US-262706P.	
XX	02-FEB-2001; 2001US-266020P.	

(INCY-) INCYTE GENOMICS INC.

Sanjanwala MM, Yao MG, Au-young J, Baughn MR, Arvizu C, Ring HZ;
Lee EA, Ding L, Hafalia AJA, Tang YT, Yue H, Tribouley CM;
Lu DAM, Lal PG, Warren BA, Yang J, Wallia NK, Nguyen DB;
Gandhi AR, Lu Y, Ison CH;
WPI: 2002-519668/55.
P-PSDB; RAE25019.

Novel human drug metabolizing polypeptide, useful in diagnosis,
prevention or treatment of autoimmune/inflammatory, cell proliferative,
neurological, developmental, endocrine, metabolic and gastrointestinal
disorders

Claim 72; Page 161-162; 169pp; English.

The invention relates to an isolated human drug metabolising enzyme (DME)
and its nucleotide. DME is useful for diagnosing, treating or preventing
disorders associated with aberrant expression of DME, where the disorders
are selected from autoimmune/inflammatory disorder such as acquired
immunodeficiency syndrome (AIDS), asthma, atherosclerosis, psoriasis,
uveitis; a cell proliferative disorder such as arteriosclerosis,
cirrhosis, hepatitis, and cancer; a neurological disorder such as
Alzheimer's disease, Huntington's disease, dementia, Parkinson's disease;
a developmental disorder such as renal tubular acidosis, epilepsy,
anemia; an endocrine disorder such as adenoma, thrombosis and
infections; an eye disorder such as conjunctivitis, glaucoma, cataract;
metabolic disorder such as cystic fibrosis, diabetes and goitre; a
gastrointestinal disorder such as anorexia, peptic ulcer; and liver
disorders. DME is useful in a number of drug screening techniques and to
analyse the proteome of a tissue or cell type. The invention is useful
for creating knock-in humanised animals or transgenic animals to model
human diseases, in somatic or germline gene therapy, to generate a
transcript image of a tissue or cell type, for detecting differences in
the chromosomal location due to translocation, inversion, etc. among
normal, carrier or affected individuals, and as hybridisation probes for
mapping naturally occurring genomic sequences. The present sequence is
human DME-4 cDNA.

Sequence 2745 BP; 704 A; 612 C; 737 G; 692 T; 0 other;

Query Match 96.1%; Score 2738.2; DB 24; Length 2745;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2737; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy	81	ATGTGGGGCCGACGGCGCGCGCGCTGCCCCGGGAACTGCGCGCGCGCGCGGAGGCG	140
Db	1	ATGTGGGGGGCGACGGCGCGCGCGCTGCCCCGGGAACTGCGCGCGCGCGCGGAGGCG	60
Qy	141	CTGTTGGTGTCTTGGCGCTACTGGCGTTGGCGGGGCTGGGCTCGGTGCTGGCGGCGCAG	200
Db	61	CTGTTGGTGTCTTGGCGCTACTGGCGTTGGCGGGGCTGGGCTCGGTGCTGGCGGCGCAG	120
Qy	201	CGTGGGGCGGGGCGGGGCTGCGAGCGCGGACCCCGCGCACCCCGCGCGCGCGGCGG	260
Db	121	CGTGGGGCGGGGCGGGGCTGCGAGCGCGGACCCCGCGCACCCCGCGCGCGGCGG	180
Qy	261	CGCAGCGGCTATGCGCGCGCGCGCGTGGCGGAAACCGCGTGGCGCGCGGGGCGAG	320
Db	181	CGCAGCGGCTATGCGCGCGCGCGCGTGGCGGAAACCGCGTGGCGCGCGGGGCGAG	240
Qy	321	CGCGTGGCGGTGACGTGCGGGCGAGAGCTGGGCTGCGAGGAGAGAGCGTGGCGGTG	380
Db	241	CGCGTGGCGGTGACGTGCGGGCGAGAGCTGGGCTGCGAGGAGAGAGCGTGGCGGTG	300
Qy	381	CACGAGATTACATCTACTCTCAGGACCGCATCTCACTGACCGCGCGCTGCCCGAGCGC	440
Db	301	CACGAGATTACATCTACTCTCAGGACCGCATCTCACTGACCGCGCGCTGCCCGAGCGC	360
Qy	441	TGGAAACCGGCTGTGCAAGAGAGAAATATGATTATGATATTTGCCAGGACATCTGTT	500
Db	361	TGGAAACCGGCTGTGCAAGAGAGAAATATGATTATGATATTTGCCAGGACATCTGTT	420

```
QY 1581 CACCAGCTGAGGCTGCATTGCTGTGGAAGCAGGAATGGATACCTTATCATCATCTC 1640
Db 1501 CACCAGCTGAGGCTGCATTGCTGTGGAAGCAGGAATGGATACCTTATCATCATCTC 1560
QY 1641 TGCAGAAAGAACTGCCAGGAGAAATCAGAAATTCATCTTGAGGAGGATGGATCTTTATTT 1700
Db 1561 TGCAGAAAGAACTGCCAGGAGAAATCAGAAATTCATCTTGAGGAGGATGGATCTTTATTT 1620
QY 1701 CACGAAAGCTCCAGAAATGTCAGGCTGCGAGGAGGAGAGTGCAGTGCAGTTTCGTT 1760
Db 1621 CACGAAAGCTCCAGAAATGTCAGGCTGCGAGGAGGAGAGTGCAGTGCAGTTTCGTT 1680
QY 1761 CCATCTTTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATG 1820
Db 1681 CCATCTTTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATG 1740
QY 1821 TTATGAAGCCCTCGTGTATCAAGGAGCCCATCGAAGGAGACTGTGAGCCAGGACTCTGCC 1880
Db 1741 TTATGAAGCCCTCGTGTATCAAGGAGCCCATCGAAGGAGACTGTGAGCCAGGACTCTGCC 1800
QY 1881 CAACAAAGACTTAGCTAAGCAGTGACAGAAACCCACCAAACTAGGCTGCAATTTGTTG 1940
Db 1801 CAACAAAGACTTAGCTAAGCAGTGACAGAAACCCACCAAACTAGGCTGCAATTTGTTG 1860
QY 1941 AAGAGGCAATCATTTGCCATTTGCAAGTGTGTTGGATTTAGTAAATGTAATAA 2000
Db 1861 AAGAGGCAATCATTTGCCATTTGCAAGTGTGTTGGATTTAGTAAATGTAATAA 1920
QY 2001 GCTTTGACTTATTTTCAGAACTTTTAAATGTTCCAAATACCCCTATTTTCAAGGGTA 2060
Db 1921 GCTTTGACTTATTTTCAGAACTTTTAAATGTTCCAAATACCCCTATTTTCAAGGGTA 1980
QY 2061 ATCTAAGATGTTAACTTGGTATTTAGAAATTTAAACCTTATAATATTTTCTATCA 2120
Db 1981 ATCTAAGATGTTAACTTGGTATTTAGAAATTTAAACCTTATAATATTTTCTATCA 2040
QY 2121 ARAWRWATTTTACAGTCTGCTCTTTTACTCTCATTTAGCAAAAAGATAAAGATTTTAT 2180
Db 2041 AGATGTATATTTTACAGTCTGCTCTTTTACTCTCATTTAGCAAAAAGATAAAGATTTTAT 2100
QY 2181 TTTGGTATTTTACAAGATTTCCAGGTACGAAGATATCTGATGGTGGAAATCAGGTTCA 2240
Db 2101 TTTGGTATTTTACAAGATTTCCAGGTACGAAGATATCTGATGGTGGAAATCAGGTTCA 2160
QY 2241 AGCAACGTACTTTGCATTAACCTGATAATACCTCAGCTGCGGGTTAAAGTTTCCCAAGTA 2300
Db 2161 AGCAACGTACTTTGCATTAACCTGATAATACCTCAGCTGCGGGTTAAAGTTTCCCAAGTA 2220
QY 2301 TAGAGAGACTGTCACTAGGAACATTTGATTTATTTAGGTCATTTGAGATCTTCTAGA 2360
Db 2221 TAGAGAGACTGTCACTAGGAACATTTGATTTATTTAGGTCATTTGAGATCTTCTAGA 2280
QY 2361 TGTATTTTAAAGAAATGCTTTTGGTTATGTTGCTTACACAGTTTAACTCCATAAT 2420
Db 2281 TGTATTTTAAAGAAATGCTTTTGGTTATGTTGCTTACACAGTTTAACTCCATAAT 2340
QY 2421 GTTCATGTCAGCCAAAGAGACTAACCAAAAGCTGAAATCTCAGAGAAATTTGCTTTAC 2480
Db 2341 GTTCATGTCAGCCAAAGAGACTAACCAAAAGCTGAAATCTCAGAGAAATTTGCTTTAC 2400
QY 2481 TAAGCTGAGTCACTTGAGAGCGAACTTTTAACTCCGACCTGTAGTGGCTGGTTC 2540
Db 2401 TAAGCTGAGTCACTTGAGAGCGAACTTTTAACTCCGACCTGTAGTGGCTGGTTC 2460
QY 2541 TACCACATGACCTTTAAACATGTTTATATCATTTTAAATTTTATGATACGCTAGTGC 2600
Db 2461 TACCACATGACCTTTAAACATGTTTATATCATTTTAAATTTTATGATACGCTAGTGC 2520
QY 2601 AGGAGAGAAATGTAATGTTTATATGAAATTTCTTTTCAAGTTTGTTCATTATAACAGT 2660
Db 2521 AGGAGAGAAATGTAATGTTTATATGAAATTTCTTTTCAAGTTTGTTCATTATAACAGT 2580
QY 2661 TATTAATTTAAATCAGCGTTAGAGTTTGTGCTGCTGCAACTGCTGTGGAATTTCTCTGA 2720
```

```
Db 2581 TATTAATTTAAATCAGCGTTAGAGTTTGTGCTGCTGCAACTGCTGTGGAATTTCTCTGA 2640
QY 2721 GTAATTCGATTTGTGTAATGATCCAGACCAACCTCGAGATTTTGTCAACTGATTAAAGT 2780
Db 2641 GTAATTCGATTTGTGTAATGATCCAGACCAACCTCGAGATTTTGTCAACTGATTAAAGT 2700
QY 2781 CAATATGAATGATTAATAAAGATGTGAGAACAAAAA 2825
Db 2701 CAATATGAATGATTAATAAAGATGTGAGAACACTGMAAAAAA 2745

RESULT 3
AAL26522/c
ID AAL26522 standard; cDNA; 2290 BP.
XX
AC AAL26522;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 18979.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 24-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer
XX
PS Claim 1; Page 3534-3535; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 2290 BP; 508 A; 676 C; 572 G; 528 T; 6 other;
Query Match 73.0%; Score 2081.8; DB 22; Length 2290;
Best Local Similarity 99.7%; Pred.No. 0;
Matches 2105; Conservative 1; Mismatches 3; Indels 2; Gaps 2;
QY 12 CCATATAACGGCGCCAGTGTGCTGGAATTCGCCCTTGCGCAGATCGCTGGCTGCAGTT 71
Db 2280 CCATATAACGGCGCCGCTGCTGGAATTCGCCCTTGCGCAGATCGCTGGCTGCAGTT 2222
QY 72 GCGCGGCGCATGTGGGGGGCGACGCGCGCGCGCTGCCCGGGGAACTGCGCGCGCGC 131
Db 2221 GCGCGGCGCATGTGGTGTGGCGCACGCGCGCGCGCTGCCCGGGGAACTGCGCGCGCGC 2162
```

Qy 132 CGGAGCGCTGTTGGTCTCCCTGGCGCTACTGSCGTTGGCCGGGCTGGGCTCGTCTG 191
Db |||||
Qy 2161 CGGAGCGCTGTTGGTCTCCCTGGCGCTACTGCGCTGTTGGCCGGGCTGGGCTCGTCTG 2102
Db |||||
Qy 192 CGGCGCAGCGTGGGCGCGGGCGCGGGCTGCGGAGCCGGGACCCCGCGCACCCCGCGC 251
Db |||||
Qy 2101 CGGCGCAGCGTGGGCGCGGGCGCGGGCTGCGGAGCCGGGACCCCGCGCACCCCGCGC 2042
Db |||||
Qy 252 CCGGGCGCGCGAGCGCGTCAATGCGCGGGCGCGCGTGGCGGCGGCGGCGGCGGCGGCGG 311
Db |||||
Qy 2041 CCGGGCGCGCGAGCGCGTCAATGCGCGGGCGCGCGTGGCGGCGGCGGCGGCGGCGGCGG 1982
Db |||||
Qy 312 CGGGCGAGCGGCGCGCTGCGAGTGCAGGCGGCGAGGCTGCGGCTGCGAGGAGGAGC 371
Db |||||
Qy 1981 CGGGCGAGGCGGCTGCGAGTGCAGGCTGCGAGGCGAGGAGCTGCGGCTGCGAGGAGGAGC 1922
Db |||||
Qy 372 GTGCGGCTGCACCAAGATTAACATCTACCTCAGCGACCGCATCTCACCTGCACCGCGCGCTG 431
Db |||||
Qy 1921 GTGCGGCTGCACCAAGATTAACATCTACCTCAGCGACCGCATCTCACCTGCACCGCGCGCTG 1862
Db |||||
Qy 432 CCGAGCGCTGGAACCCCGCTGTGCAAGAGAGAAATATGATTAATGATTAATTTGCCCGAG 491
Db |||||
Qy 1861 CCGAGCGCTGGAACCCCGCTGTGCAAGAGAGAAATATGATTAATGATTAATTTGCCCGAG 1802
Db |||||
Qy 492 ACATCTGTTATCATAGCATTTTATAATGAGCCCTGCTCACTCTCTCGGACAGTTTAC 551
Db |||||
Qy 1801 ACATCTGTTATCATAGCATTTTATAATGAGCCCTGCTCACTCTCTCGGACAGTTTAC 1742
Db |||||
Qy 552 AGTGTCTTTGAGACATCCCGGATATCTGCTAGAAGAGTGTCTTTGAGAGTACTAC 611
Db |||||
Qy 1741 AGTGTCTTTGAGACATCCCGGATATCTGCTAGAAGAGTGTCTTTGAGAGTACTAC 1682
Db |||||
Qy 612 AGTGATGAGAGCACCTGGAAGAGCGCTTGCCCAATGAGCTTTCCGAGCTGCCCAAGGTG 671
Db |||||
Qy 1681 AGTGATGAGAGCACCTGGAAGAGCGCTTGCCCAATGAGCTTTCCGAGCTGCCCAAGGTG 1622
Db |||||
Qy 672 CGCTGTATCCGCGCAACAGAGAGAGCGCTGCTGCGAGCGCGGCTGCTGGGCGGCTCT 731
Db |||||
Qy 1621 CGCTGTATCCGCGCAACAGAGAGAGCGCTGCTGCGAGCGCGGCTGCTGGGCGGCTCT 1562
Db |||||
Qy 732 CGCGCGAGGGCGGATGTTCTGACCTTCTGACCTGTCTCACTGTGAGTGCACAGAGGGTGG 791
Db |||||
Qy 1561 CGCGCGAGGGCGGATGTTCTGACCTTCTGACCTGTCTCACTGTGAGTGCACAGAGGGTGG 1502
Db |||||
Qy 792 CTGAGCGCGCTGTGCGAGAGATCCATGAAGAGAGTCCGAGTGGTGTGCCCGGTGATT 851
Db |||||
Qy 1501 CTGAGCGCGCTGTGCGAGAGATCCATGAAGAGAGTCCGAGTGGTGTGCCCGGTGATT 1442
Db |||||
Qy 852 GATGTGATCGACTGGAACACCTTCGAATACCTGGGGAACCTCCGGGAGCCCGACATCGGC 911
Db |||||
Qy 1441 GATGTGATCGACTGGAACACCTTCGAATACCTGGGGAACCTCCGGGAGCCCGACATCGGC 1382
Db |||||
Qy 912 GGTTCGACTCGAGCGTGTGTTCACTGTGGGACACAGTTCCTGAGAGGAGGATACGG 971
Db |||||
Qy 1381 GGTTCGACTCGAGCGTGTGTTCACTGTGGGACACAGTTCCTGAGAGGAGGATACGG 1322
Db |||||
Qy 972 ATGCAATCCCGCTCGATGTCTCAGGTCTCCAAATGCGTGTGGGCTGTTGCTGTG 1031
Db |||||
Qy 1321 ATGCAATCCCGCTCGATGTCTCAGGTCTCCAAATGCGTGTGGGCTGTTGCTGTG 1262
Db |||||
Qy 1032 AGTAAGAAATATTTGAAATATCTGGGCTTATGATACAGGAATGGAAGTTGGGAGGA 1091
Db |||||
Qy 1261 AGTAAGAAATATTTGAAATATCTGGGCTTATGATACAGGAATGGAAGTTGGGAGGA 1202
Db |||||
Qy 1092 GAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTGGGGTCTTGAAACACACCA 1151
Db |||||
Qy 1201 GAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTGGGGTCTTGAAACACACCA 1142
Db |||||
Qy 1152 TGTTCCCATGTGGCCATGTTTTCCTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTG 1211
Db |||||
Qy 1141 TGTTCCCATGTGGCCATGTTTTCCTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTG 1082
Db |||||

Qy 1212 GCCAACAGTGTTCGTGCTGAGCTGAAGTATGATGATGAATTTAAAGAGCTCTACTACCAT 1271
Db |||||
Qy 1081 GCCAACAGTGTTCGTGCTGAGCTGAAGTATGATGATGAATTTAAAGAGCTCTACTACCAT 1023
Db |||||
Qy 1272 CGCAACCCCGCTGCGCGCTTGGAACTTTTGGGGATGTACAGAGAGGAGAGCTCCGG 1331
Db |||||
Qy 1022 CGCAACCCCGCTGCGCGCTTGGAACTTTTGGGGATGTGACAGAGAGGAGAGCTCCGG 963
Db |||||
Qy 1332 GACAACTCCAGTGAAGACTTCAAGTGTTCCTGAGAGCTGTGTATCCAGAACTGTCAT 1391
Db |||||
Qy 962 GACAACTCCAGTGAAGACTTCAAGTGTTCCTGAGAGCTGTGTATCCAGAACTGTCAT 903
Db |||||
Qy 1392 GTCCCTGAGGACAGGCGCTGCTTCTTCCGGATGCTCCAGAAACAAAGGACTTAAACAGACTAC 1451
Db |||||
Qy 902 GTCCCTGAGGACAGGCGCTGCTTCTTCCGGATGCTCCAGAAACAAAGGACTTAAACAGACTAC 843
Db |||||
Qy 1452 TGCTTTGACTATTAACCTCCGATGAAACCCAGATTTGTGGGACACAGGCTCATTTCTGTAC 1511
Db |||||
Qy 842 TGCTTTGACTATTAACCTCCGATGAAACCCAGATTTGTGGGACACAGGCTCATTTCTGTAC 783
Db |||||
Qy 1512 CTCTGTATGGGATGGGCGCAGATCAAGTTCCTGAGTACAGCTCCAGAAAGAAATACGC 1571
Db |||||
Qy 782 CTCTGTATGGGATGGGCGCAGATCAAGTTCCTGAGTACAGCTCCAGAAAGAAATACGC 723
Db |||||
Qy 1572 TATAACACCCAGGCTGAGGCTGCTTGTGTGGAAGAGAAATGGAATACCTTTATC 1631
Db |||||
Qy 722 TATAACACCCAGGCTGAGGCTGCTTGTGTGGAAGAGAAATGGAATACCTTTATC 663
Db |||||
Qy 1632 ATGCATCTCTGCGAAGAACTGCCCGAGAGAAATCAGAAGTTCATCTTGCAGGAGGATGGA 1691
Db |||||
Qy 662 ATGCATCTCTGCGAAGAACTGCCCGAGAGAAATCAGAAGTTCATCTTGCAGGAGGATGGA 603
Db |||||
Qy 1692 TCTTTATTTACCAACAGTCCAGAAATGTCTCAGGCTGCGAGGAGGAGTCCAGTGAC 1751
Db |||||
Qy 602 TCTTTATTTACCAACAGTCCAGAAATGTCTCAGGCTGCGAGGAGGAGTCCAGTGAC 543
Db |||||
Qy 1752 AGTTTCGTTTCCACTCTTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAA 1811
Db |||||
Qy 542 AGTTTCGTTTCCACTCTTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAA 483
Db |||||
Qy 1812 GAGCGATGTTTGAAGCGCTGCTGTATCAAGAGGCCCATCGAAGGAGACTGTGAGGCCAG 1871
Db |||||
Qy 482 GAGCGATGTTTGAAGCGCTGCTGTATCAAGAGGCCCATCGAAGGAGACTGTGAGGCCAG 423
Db |||||
Qy 1872 GACTCTGCCCAACAAAGACTTAGCTAAGCAGTGCACAGAACCCACCAAACTAGGCTGC 1931
Db |||||
Qy 422 GACTCTGCCCAACAAAGACTTAGCTAAGCAGTGCACAGAACCCACCAAACTAGGCTGC 363
Db |||||
Qy 1932 ATTGCTTTGAAGAGGCAATCATTTTGCCTTTGTGAAAGTGTGTGGATTTAGTAAAAA 1991
Db |||||
Qy 362 ATTGCTTTGAAGAGGCAATCATTTTGCCTTTGTGAAAGTGTGTGGATTTAGTAAAAA 303
Db |||||
Qy 1992 TGTGAATAAGCTTTGACTTATTTTGAAGACTTTTAAATGTTTCCAAAATACCTATTTT 2051
Db |||||
Qy 302 TGTGAATAAGCTTTGACTTATTTTGAAGACTTTTAAATGTTTCCAAAATACCTATTTT 243
Db |||||
Qy 2052 CAAGGGTATCTGAGATGTTTAACTTGGTATTTAGAAATTAACCTTATTAATTT 2111
Db |||||
Qy 242 CAAGGGTATCTGAGATGTTTAACTTGGTATTTAGAAATTAACCTTATTAATTT 183
Db |||||
Qy 2112 TTTCTATCAAR 2122
Db |||||
Qy 182 TTTCTATCCAA 172
Db |||||

RESULT 4

AAL26538/c

ID AAL26538 standard; cDNA; 2290 BP.

XX

AC AAL26538;

XX

DT 07-DEC-2001 (first entry)

XX

Qy 1512 CTCTGTCATGGATGGGCGAGAAATCAGTTTTCAGTAGTACAGTCCCGAGAAAGAAATACGC 1571
Db |||||||
Qy 782 CTCTGTCATGGATGGGCGAGAAATCAGTTTTCAGTAGTACAGTCCCGAGAAAGAAATACGC 723
Db |||||||
Qy 1572 TATTAACACCCACACAGCTGAGGGCTGATTTCTGTGTGGAAGCAGGAATGATACCTTATC 1631
Db |||||||
Qy 722 TATAACACCCACACAGCTGAGGGCTGATTTCTGTGTGGAAGCAGGAATGATACCTTATC 663
Db |||||||
Qy 1632 ATGATCTCTCTCGGAGAACTGCCCGAGAGATCAGAGATCAGAGTTTCATCTTGCAGGAGATGGA 1691
Db |||||||
Qy 662 ATGATCTCTCTCGGAGAACTGCCCGAGAGATCAGAGATCAGAGTTTCATCTTGCAGGAGATGGA 603
Db |||||||
Qy 1692 TCTTTATTTACGACAGCTCCAGAAATGTCTCAGGCTGCGAGAGAGAGTCCAGTGCAC 1751
Db |||||||
Qy 602 TCTTTATTTACGACAGCTCCAGAAATGTCTCAGGCTGCGAGAGAGAGTCCAGTGCAC 543
Db |||||||
Qy 1752 AGTTTGTCTCCTTACGAGACTGACACCACTCGGATCATCAGAAATGTTCTTCAAA 1811
Db |||||||
Qy 542 AGTTTGTCTCCTTACGAGACTGACACCACTCGGATCATCAGAAATGTTCTTCAAA 483
Db |||||||
Qy 1812 GAGCGCATGTTAGAGCTCTGATATCAAGGAGCCCATCGAAGGAGACTGTGAGCCAG 1871
Db |||||||
Qy 482 GAGCGCATGTTAGAGCTCTGATATCAAGGAGCCCATCGAAGGAGACTGTGAGCCAG 423
Db |||||||
Qy 1872 GACTCTGCCCCAACAAAGACTTAGCTTAAGCAGTGCACCAACCCCAAACTAGGCTGC 1931
Db |||||||
Qy 422 GACTCTGCCCCAACAAAGACTTAGCTTAAGCAGTGCACCAACCCCAAACTAGGCTGC 363
Db |||||||
Qy 1932 ATTGCTTTGAGAGGCAATCATTTTGGCCATTTGTTGAAAGTTGTTGTTAGTAAAAA 1991
Db |||||||
Qy 362 ATTGCTTTGAGAGGCAATCATTTTGGCCATTTGTTGAAAGTTGTTGTTAGTAAAAA 303
Db |||||||
Qy 1992 TGTGAATAAGCTTTGACTTATTTTGAACTTTTAAATGTTTCCAAATACCTTATTTT 2051
Db |||||||
Qy 302 TGTGAATAAGCTTTGACTTATTTTGAACTTTTAAATGTTTCCAAATACCTTATTTT 243
Db |||||||
Qy 2052 CAAGGGTAAATCGTAAGATGTTAAACCTTTGTTATTTAGAAAATTTAAACCTTATTAATTT 2111
Db |||||||
Qy 242 CAAGGGTAAATCGTAAGATGTTAAACCTTTGTTATTTAGAAAATTTAAACCTTATTAATTT 183
Db |||||||
Qy 2112 TTTCTATCAAR 2122
Db |||||||
Qy 182 TTTCTATCCAA 172
Db |||||||

RESULT 5
AAC98209
ID AAC98209 standard; cDNA; 1259 BP.

AC AAC98209;

DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen nucleotide sequence SEQ ID NO:219.

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antineoplastic; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; ss.

OS Homo sapiens.

FN WO200055351-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05883.

PR 12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WFI; 2000-587534/55.

P-PSDB; AAB53452.

Colon cancer associated gene sequences, referred to as colon cancer
antigens, useful for the treatment, prevention, and diagnosis of colon
disorders such as colon cancer -

Claim 1; Page 642; 2104pp; English.

AAC987991 to AAC98763 encode the human colon cancer associated proteins,
called human colon cancer antigens, given in AAB53234 to AAB54006. The
human colon cancer antigens can have cytostatic, cardioactive, muscular;
neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and
vulnary, nephrotropic, antineoplastic and antibacterial activities, and
can be used in gene therapy. The colon cancer antigen polynucleotides,
proteins and antibodies to the proteins are useful for the prevention,
treatment and diagnosis of colon disorders, such as colon cancer. The
polynucleotides may be used in diagnostics and research, such as for
chromosome identification, and as hybridization probes. The proteins
may also be used to prevent diseases such as neural disorders, immune
system disorders, muscular disorders, reproductive disorders,
gastrointestinal disorders, wounds, renal disorders, infectious
diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
AAB54007 represent sequences used in the exemplification of the present
invention.

Sequence 1259 BP; 284 A; 310 C; 383 G; 278 T; 4 other;

Query Match 42.4%; Score 1208.4; DB 21; Length 1259;

Best Local Similarity 98.4%; Pred. No. 2.7e-210;

Matches 1226; Conservative 4; Mismatches 15; Indels 1; Gaps 1;

Qy 292 CGGCGAACCGCGCTGGGCGCGCGGGGCGAGCGGTGCGGCTGCGAGCTCAGGCGGAGGAGC 351

Db 7 CGGTGCGGGAACCGCTGGGCGGGGAGCGAGCGGTGCGGCTGCGAGCTCAGGCGGAGGAGC 66

Qy 352 TGGGCTGTCAGGAGGAGAGCGTGGCGCTGCAACAGATTAACATCTACCTCAGCGACCGCA 411

Db 67 TGGGCTGTCAGGAGGAGAGCGTGGCGCTGCAACAGATTAACATCTACCTCAGCGACCGCA 126

Qy 412 TCTCAGTGCACCGCGCTGCGGCGCGCGGGGCGAGCGGTGCGGCTGCGAGCTCAGGCGGAGGAGC 471

Db 127 TCTCAGTGCACCGCGCTGCGGCGCGCGGGGCGAGCGGTGCGGCTGCGAGCTCAGGCGGAGGAGC 186

Qy 472 ATTATGATAATTTGCCCGAGGACATCTGTTATCATAGCATTTTATAATGAAGCCTGGTCAA 531

Db 187 ATTATGATAATTTGCCCGAGGACATCTGTTATCATAGCATTTTATAATGAAGCCTGGTCAA 246

Qy 532 CTCTCTTTCGAGCAGTTTACAGTGTCTTGGAGACATCCCCGGATATCTCTGTAGAGAGAG 591

Db 247 CTCTCTTTCGAGCAGTTTACAGTGTCTTGGAGACATCCCCGGATATCTCTGTAGAGAGAG 306

Qy 592 TGTATCTTGTAGATGACTACAGTATAGAGACCATGAGAGGCGCTTGGCCCAATCAGC 651

Db 307 TGTATCTTGTAGATGACTACAGTATAGAGACCATGAGAGGCGCTTGGCCCAATCAGC 366

Qy 652 TTTTCGGGACTGCCCAAGTGGCTGATCGCGCGCCAAACAAGAGAGAGGCGCTGTGGGAG 711

Db 367 TTTTCGGGACTGCCCAAGTGGCTGATCGCGCGCCAAACAAGAGAGAGGCGCTGTGGGAG 426

Qy 712 CCGGCTGTGGGGCGCTCTGCGGCGAGGGCGGATGTTCTGACCTTCTGAGCTGTCACT 771

Db 427 CCGGCTGTGGGGCGCTCTGCGGCGAGGGCGGATGTTCTGACCTTCTGAGCTGTCACT 486

Qy 772 GTGAGTCCCAAGAGGCTGTGGAGCGCTGCTGTCAGAGGATCCATGAGAGAGGAGTCGG 831

Db 487 GTGAGTCCCAAGAGGCTGTGGAGCGCTGCTGTCAGAGGATCCATGAGAGAGGAGTCGG 545

Qy 832 CAGTGTGTGCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 891

Db 546 CAGTGGTGTGCCCCGGTGTATGATGTSATCGACTGGAACACACCTTCGAATACCTCGGGAACT 605
Qy 892 CCGGGAGCCCCAGATCGGCGGTTTCGACTGGAGGCTGGTGTTCACGTGGCACACAGTTC 951
Db 606 CCGGGAGCCCCAGATCGGCGGTTTCGACTGGAGGCTGGTGTTCACGTGGCACACAGTTC 665
Qy 952 CTGAGAGGGAGAGATACGATGCAATCCCGCTCGATGTCATCAGGTCTCCAAATGG 1011
Db 666 CTGAGAGGGAGAGATACGATGCAATCCCGCTCGATGTCATCAGGTCTCCAAATGG 725
Qy 1012 CTGTGGGCTGTTTGTCTGTAGTAAAGAAATATTTTGAATATCTGGGCTTATGATACAG 1071
Db 726 CTGTGGGCTGTTTGTCTGTAGTAAAGAAATATTTTGAATATCTGGGCTTATGATACAG 785
Qy 1072 GAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTG 1131
Db 786 GAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTG 845
Qy 1132 GGGTCTTGGAACACACCCATGTTCCCATGTTGCCATGTTTCCCAAGCAAGCTCCCT 1191
Db 846 GGGTCTTGGAACACACCCATGTTCCCATGTTGCCATGTTTCCCAAGCAAGCTCCCT 905
Qy 1192 ACTCCGCAACAGGCTCTGSCCAACAGTGTTCGTGAGCTGAAGTATGGATGATGAAT 1251
Db 906 ACTCCGCAACAGGCTCTGSCCAACAGTGTTCGTGAGCTGAAGTATGGATGATGAAT 965
Qy 1252 TTAAGAGCTCTACTACCATCGCAACCCCGCTGCGCTTGGAACTTTTGGGATGTGA 1311
Db 966 TTAAGAGCTCTACTACCATCGCAACCCCGCTGCGCTTGGAACTTTTGGGATGTGA 1025
Qy 1312 CAGAGAGAGAGCTCGGGAACAGTCTCCAGTGTAAAGACTTCAAGTGTTCCTGGAGA 1371
Db 1026 CAGAGAGAGAGCTCGGGAACAGTCTCCAGTGTAAAGACTTCAAGTGTTCCTGGAGA 1085
Qy 1372 CTGTGATCCAGACTCGATGTCCTGAGGACAGGCTGGTCTTCTGGGATGCTCCAGA 1431
Db 1086 CTGTGATCCAGAACTCGATGTCCTGAGGACAGGCTGGTCTTCTGGGATGCTCCAGA 1445
Qy 1432 ACAAGGACTTAAACAGACTACTGCTTGAATATAACCTCCGATGAAACACAGATTTGG 1491
Db 1146 ACAAGGACTTAAACAGACTACTGCTTGAATATAACCTCCGATGAAACACAGATTTGG 1205
Qy 1492 GACACAGGCTATCTGTACTCTGTATGATGGATGGGCCAGAAATCA 1537
Db 1206 GACACAGGCTATCTGTACTCTGTATGATGGATGGGCCAGAAATCA 1251

RESULT 6

ABV23465/c
ID ABV23465 standard; cDNA; 1352 BP.

XX AC ABV23465;

XX 16-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 23456.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX pharmacodynamic marker; gene; ss.

XX Homo sapiens.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US05171.

XX 17-FEB-2000; 2000US-183319P.

XX 16-MAR-2000; 2000US-189862P.

XX 25-MAY-2000; 2000US-207454P.

XX 09-JUN-2000; 2000US-211314P.

PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

PS Claim 1; Page 4286-4287; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient;

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 1352 BP; 340 A; 335 C; 299 G; 368 T; 10 other;

Qy Query Match 39.1%; Score 1115.2; DB 23; Length 1352;

Db Best Local Similarity 99.5%; Pred. No. 2.4e-193;

Qy Matches 1128; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Db 989 TGTTCATCAGTCTCCAAATGCGTGGGCTGTTTGTCTGTGAGTAAAGAAATTTTGA 1048

Qy 1304 TGCCATGTGCTCTCCAAATGCGTGGGCTGTTTGTCTGTGAGTAAAGAAATTTTGA 1245

Qy 1049 ATATCTGGGCTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTC 1108

Db 1244 ATATCTGGGCTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTC 1185

Qy 1109 CTTTAGGATCTGGCAGTGTGCTGGGTTCTGGAACACACCCATGTTCCCATGTTGGCCA 1168

Db 1184 CTTTAGGATCTGGCAGTGTGCTGGGTTCTGGAACACACCCATGTTCCCATGTTGGCCA 1125

Qy 1169 TGTTCATCAGTCTCCAAAGTCTCCCTACTCCCGCAACAGGCTCTGGCCAACAGTGTTCGTCG 1228

Db 1124 TGTTCATCAGTCTCCAAAGTCTCCCTACTCCCGCAACAGGCTCTGGCCAACAGTGTTCGTCG 1066

Qy 1229 AGCTGAAGTATGGATGGAATTTTAAAGAGTCTACTACATCGCAACCCCGTGGCCG 1288

Db 1065 AGCTGAAGTATGGATGGAATTTTAAAGAGTCTACTACATCGCAACCCCGTGGCCG 1006

Qy 1289 CTTTGGACCTTTTGGGATGTGACAGAGAGGAGGAGCTCCGGACAGCTCCAGTGTAA 1348

Db 1005 CTTTGGACCTTTTGGGATGTGACAGAGAGGAGGAGCTCCGGACAGCTCCAGTGTAA 946

Qy 1349 AGACTTCAAGTGTTCCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCC 1408

Db 945 AGACTTCAAGTGTTCCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCC 886

Qy 1409 TGGCTTCTTCGGGATGCTCCAGAAACAAAGGACTAAACAGACTA CTGCTTTGACTATAACCC 1468

Db 885 TGGCTTCTTCGGGATGCTCCAGAAACAAAGGACTAAACAGACTA CTGCTTTGACTATAACCC 826

Qy 1469 TCCGATGAAACCAAGTGTGGGACACAGGCTCATCTCTACCTCTCTGCTATGGATGGG 1528

Db 825 TCCGATGAAACCAAGTGTGGGACACAGGCTCATCTCTACCTCTCTGCTATGGATGGG 766

Qy 1529 CCAGATCAGTGTCTTTCGAGTACAGCTCCAGAGAAATAGCTATACACCCACAGCC 1588
Db 765 CCAGATCAGTGTCTTTCGAGTACAGCTCCAGAGAAATAGCTATACACCCACAGCC 706
Qy 1589 TGAGGGTGCATTCGTGTGGAAGCAGGAATGGATACCTTATCATGATCTCTCGAAGA 1648
Db 705 TGAGGGTGCATTCGTGTGGAAGCAGGAATGGATACCTTATCATGATCTCTCGAAGA 646
Qy 1649 AACTGCCCCAGAGATCAGAGTTCATCTTGCAGGAGATGGATCTTATTTACAGACA 1708
Db 645 AACTGCCCCAGAGATCAGAGTTCATCTTGCAGGAGATGGATCTTATTTACAGACA 586
Qy 1709 GTCCAGAAATGTCTCAGGCTGCAGGAGAGAGTCTCAGTACACAGTTCGTTCCACTCTT 1768
Db 585 GTCCAGAAATGTCTCAGGCTGCAGGAGAGAGTCTCAGTACACAGTTCGTTCCACTCTT 526
Qy 1769 ACGAGCTGCACCACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGAAG 1828
Db 525 ACGAGCTGCACCACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGAAG 466
Qy 1829 CCTCGTGTATCAAGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAG 1888
Db 465 CCTCGTGTATCAAGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAG 406
Qy 1889 ACTTAGCTAAGCAGTGCAGAACCCACCAAACTAGGCTGCATTCGTTTGAAGAGGCA 1948
Db 405 ACTTAGCTAAGCAGTGCAGAACCCACCAAACTAGGCTGCATTCGTTTGAAGAGGCA 346
Qy 1949 ATCAATTTGCAATTTGAAAGTGTGTGATTTAGTAAATGTAAGTAAAGTAAAGTAAAG 2008
Db 345 ATCAATTTGCAATTTGAAAGTGTGTGATTTAGTAAATGTAAGTAAAGTAAAGTAAAG 286
Qy 2009 CTTATTTTGAAACTTTTAAATGTTCCAAATACCTTATTTTCAAGGGTAAATCGTAAG 2068
Db 285 CTTATTTTGAAACTTTTAAATGTTCCAAATACCTTATTTTCAAGGGTAAATCGTAAG 226
Qy 2069 ATGTTAACCTTGGTATTTAGAAATTTAAACCTTATTAATATTTTCTATCAAR 2122
Db 225 ATGTTAACCTTGGTATTTAGAAATTTAAACCTTATTAATATTTTCTATCAAR 172

RESULT 7

ABV29323/c

ID ABV29323 standard; cDNA; 1352 BP.

XX AC ABV29323;

XX DT 16-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 29314.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WO200160860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US05171.

XX PR 17-FEB-2000; 2000US-183319P.

XX PR 16-MAR-2000; 2000US-189862P.

XX PR 25-MAY-2000; 2000US-207454P.

XX PR 09-JUN-2000; 2000US-211314P.

XX PR 18-JUL-2000; 2000US-219007P.

XX PR 13-DEC-2000; 2000US-255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX

DR WP1; 2001-662795/76.
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX Claim 1; Page 6275; 11750pp; English.
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX

SQ Sequence 1352 BP; 340 A; 335 C; 299 G; 368 T; 10 other;

Query Match 39.1%; Score 1115.2; DB 23; Length 1352;

Best Local Similarity 99.5%; Pred. No. 2.4e-193;

Matches 1128; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Qy 989 TGTATCATCAGTCTCCAAACAAATGCTGTGGGCTGTTTGTGTGTGAGTAAGAAATATTGA 1048
Db 1304 TGCCATGTGTCTCCAAACAAATGCTGTGGGCTGTTTGTGTGTGAGTAAGAAATATTGA 1245
Qy 1049 ATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTC 1108
Db 1244 ATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTC 1185
Qy 1109 CTTTAGATCTGGCAGTGTGGGGTCTTGGAACACACACCCATGTTCCCATGTTGGCCA 1168
Db 1184 CTTTAGATCTGGCAGTGTGGGGTCTTGGAACACACACCCATGTTCCCATGTTGGCCA 1125
Qy 1169 TGTTTTCCCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGTGTTCGTGC 1228
Db 1124 TGTTTTCCCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGTGTTCGTGC 1066
Qy 1229 AGCTGAAGTATGGATGGAATTTAAAGAGTCTTACTACATCGCAACCCCGTGGCCG 1288
Db 1065 AGCTGAAGTATGGATGGAATTTAAAGAGTCTTACTACATCGCAACCCCGTGGCCG 1006
Qy 1289 CTTTGGACCTTTTGGGATGTGACAGAGAGAGAGAGCTCCGGGACAGCTCCAGTGTAA 1348
Db 1005 CTTTGGACCTTTTGGGATGTGACAGAGAGAGAGAGCTCCGGGACAGCTCCAGTGTAA 946
Qy 1349 AGACTTCAAGTGTGTTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTCGAGGACAGCC 1408
Db 945 AGACTTCAAGTGTGTTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTCGAGGACAGCC 886
Qy 1409 TGGCTTCTTCGGGATGTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATAACC 1468
Db 885 TGGCTTCTTCGGGATGTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATAACC 826
Qy 1469 TCCCGATGAACACAGATTTGGGACACAGGCTCATTTCTGCTCTGTCATGGGATGGG 1528
Db 825 TCCCGATGAACACAGATTTGGGACACAGGCTCATTTCTGCTCTGTCATGGGATGGG 766
Qy 1529 CCAGAAATCAGTGTTCGAGTACAGCTCCAGAAAGAAATACGCTATTAACACCCACAGCC 1588
Db 765 CCAGAAATCAGTGTTCGAGTACAGCTCCAGAAAGAAATACGCTATTAACACCCACAGCC 706
Qy 1589 TGAGGGTGCATTCGTGTGAAAGCAGGAATGGATACCTTATCATGATCTCTCGGAAGA 1648
Db 705 TGAGGGTGCATTCGTGTGAAAGCAGGAATGGATACCTTATCATGATCTCTCGGAAGA 646

QY 1649 AACTGCCCCAGAGAAATCAGAAATTCATCTCTGCGAGGAGATGGATCTTTATTTACGAAACA 1708
Dd 645 AACTGCCCCAGAGAAATCAGAAATTCATCTCTGCGAGGAGATGGATCTTTATTTACGAAACA 586
QY 1709 GTCCCAAGAAATGTCTCAGGCTCGAGGAGGAGTCCAGTGCACAGTTTCCTCCACTCTT 1768
Dd 585 GTCCCAAGAAATGTCTCAGGCTCGAGGAGGAGTCCAGTGCACAGTTTCCTCCACTCTT 526
QY 1769 ACAGAGACTGCACCAACTCGGATCATCAGAAATGTTCTTCAAGAGGCGCATGTTATGAAG 1828
Dd 525 ACAGAGACTGCACCAACTCGGATCATCAGAAATGTTCTTCAAGAGGCGCATGTTATGAAG 466
QY 1829 CTTGCTGTATCAAGAGGCCCATCGAAGGAGACTGTGAGGAGGAGTCTGCCCCAACAAAG 1888
Dd 465 CTTGCTGTATCAAGAGGCCCATCGAAGGAGACTGTGAGGAGGAGTCTGCCCCAACAAAG 406
QY 1889 ACTTAGCTAAGCAGTACCAAGACCCACCAAACTAGGCTGCATTCCTTGAAGAGGCA 1948
Dd 405 ACTTAGCTAAGCAGTACCAAGACCCACCAAACTAGGCTGCATTCCTTGAAGAGGCA 346
QY 1949 ATCAATTTTGCATTTGTGAAAGTTGTGATTTAGTAAATGTAATAAGCTTTGTA 2008
Dd 345 ATCAATTTTGCATTTGTGAAAGTTGTGATTTAGTAAATGTAATAAGCTTTGTA 286
QY 2009 CTTATTTTGAGAACTTTTAAATGTTCCAAAATACCCCTATTTTCAAAAGGTAATCGTAAG 2068
Dd 285 CTTATTTTGAGAACTTTTAAATGTTCCAAAATACCCCTATTTTCAAAAGGTAATCGTAAG 226
QY 2069 ATGTTAACCTCTGATTTAGAAATTAATAACCTTATATATTTTCTATCAAR 2122
Dd 225 ATGTTAACCTCTGATTTAGAAATTAATAACCTTATATATTTTCTATCAAR 172

RESULT 8

AAH33532
ID AAH33532 standard; cDNA; 1517 BP.
XX AC AAH33532;
DT 03-SEP-2001 (first entry)
XX DE Human colon cancer antigen encoding cDNA SEQ ID NO:588.
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX OS Homo sapiens.
XX PN WO20012920-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US26524.
XX PR 29-SEP-1999; 99US-0157137.
XX PR 03-NOV-1999; 99US-0163280.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX DR WPI; 2001-235357/24.
XX DR P-PSDB; AAG74101.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 1; Page 2640; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon

CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX SQ Sequence 1517 BP; 366 A; 360 C; 363 G; 425 T; 3 other;

Query Match 35.5%; Score 1012.2; DB 22; Length 1517;
Best Local Similarity 99.3%; Pred. No. 1.3e-174;
Matches 1008; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GATCTGGCAGTGTGGTGGGTTCTTGGAAACACACCCATGTTCCATGTTGGCCATGTTT 1174
Dd 479 GATCTGGCAGTGTGGTGGGTTCTTGGAAACACACCCATGTTCCATGTTGGCCATGTTT 538
QY 1175 CCCCAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGTCTTGTGAGCTGA 1234
Dd 539 CCCCAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGTCTTGTGAGCTGA 598
QY 1235 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGA 1294
Dd 599 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGA 658
QY 1295 ACCTTTTGGGATGTGACAGAGAGAGAGAGCTCCGGGCAAGCTCCAGTGTAAAGACTT 1354
Dd 659 ACCTTTTGGGATGTGACAGAGAGAGAGAGCTCCGGGCAAGCTCCAGTGTAAAGACTT 718
QY 1355 CAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCTGTGAGTGTGAGGAGAGGCTGGCTT 1414
Dd 719 CAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCTGTGAGTGTGAGGAGAGGCTGGCTT 778
QY 1415 CTTTCGGGATGCTCCAGAAACAAAGAGCTAAACAGACTACTGCTTTGACTATTAACCTCCCGA 1474
Dd 779 CTTTCGGGATGCTCCAGAAACAAAGAGCTAAACAGACTACTGCTTTGACTATTAACCTCCCGA 838
QY 1475 TGAACACAGATTTGTGGGACACAGGTCATCTGTACCTCTGTCTGATGGGATGGGCCAGAA 1534
Dd 839 TGAACACAGATTTGTGGGACACAGGTCATCTGTACCTCTGTCTGATGGGATGGGCCAGAA 898
QY 1535 TCAGTTTTTTCAGTACACGTCCTCCAGAAAGAAATACGCTATAACACCCAGCTGAGGG 1594
Dd 899 TCAGTTTTTTCAGTACACGTCCTCCAGAAAGAAATACGCTATAACACCCAGCTGAGGG 958
QY 1595 CTGCATTGCTGTGGAAGCAGGAATGATACCTTTATCATGATCTCTGCGAAGAACTGC 1654
Dd 959 CTGCATTGCTGTGGAAGCAGGAATGATACCTTTATCATGATCTCTGCGAAGAACTGC 1018
QY 1655 CCAGAGAAATCAGAAAGTTTCACTTTGAGGAGATGATCTTTATTTTCAAGAACTGCT 1714
Dd 1019 CCAGAGAAATCAGAAAGTTTCACTTTGAGGAGATGATCTTTATTTTCAAGAACTGCT 1078
QY 1715 GAAATGTCTCCAGGCTGCGAGGAGAGTGCAGTGTGATGTTTCTGTTCCACTCTTACGAG 1774
Dd 1079 GAAATGTCTCCAGGCTGCGAGGAGAGTGCAGTGTGATGTTTCTGTTCCACTCTTACGAG 1138
QY 1775 CTGCACCAACTCGGATCATCAGAAATGTTTCTTCAAGAGCGCATGTTTATGAAGCTCGT 1834
Dd 1139 CTGCACCAACTCGGATCATCAGAAATGTTTCTTCAAGAGCGCATGTTTATGAAGCTCGT 1198
QY 1835 GTATCAAGAGGCCCATCGAAGGAGACTGTGAGGAGGAGTCTGCCCCAACAAAGACTTAG 1894

1199 GTATCAAGGACCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAG 1258
1895 CTAGCAGTACACGACACCCACCAAAACCTAGGCTGCTTGGAGGCAATCATT 1954
1259 CTAGCAGTACACGACACCCACCAAAACCTAGGCTGCTTGGAGGCAATCATT 1318
1955 TTGCAATTTGTGAAGTTGTGTTGGATTAGTAAATAATGCAATAGCTTTGTACTTATT 2014
1319 TTGCCATTTGTGAAGTTGTGTTGGATTAGTAAATAATGCAATAGCTTTGTACTTATT 1378
2015 TTGAGAACTTTTAAATGTTCCAAATAACCTATTTTCAAAGGGTAATCGTAAGATGTTA 2074
1379 TTGAGAACTTTTAAATGTTCCAAATAACCTATTTTCAAAGGGTAATCGTAAGATGTTA 1438
2075 ACCCTTGGTATTAGAAATTTAAACCTTATATATTTTCTATCAARARWAWA 2129
1439 ACCCTTGGTATTAGAAATTTAAACCTTATATATTTTCTATCAARARWAWA 1493

RESULT 9

ABL90290
ID ABL90290 standard; cDNA; 1517 BP.
XX
AC ABL90290;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 852.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; gene; ss.
XX

Homo sapiens.

OS
XX
PN WO200190304-A2.
XX

29-NOV-2001.

18-MAY-2001; 2001WO-US16450.

19-MAY-2000; 2000US-205515P.

(HUMA-) HUMAN GENOME SCI INC.

Birse CE, Rosen CA;

WPI; 2002-122018/16.

P-PSDB; ABB89881.

Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders -

Claim 4; SEQ ID NO 852; 2081pp + Sequence Listing; English.

The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB89040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 1517 BP; 366 A; 360 C; 363 G; 425 T; 3 other;

Query Match 35.5%; Score 1012.2; DB 24; Length 1517;
Best Local Similarity 99.3%; Pred. No. 1.3e-174;
Matches 1008; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1115 GATCTGGCAGTGTGGGGTCTTGAAACACACCCATGTTCCCATGTGGCCATGTTTT 1174

Db 479 GATCTGGCAGTGTGGGGTCTTGAAACACACCCATGTTCCCATGTGGCCATGTTTT 538

QY 1175 CCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCACAGTGTGTCGACGCTGA 1234

Db 539 CCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCACAGTGTGTCGACGCTGA 598

QY 1235 AGTATGGATGGATGAATTTAAAGAGCTTACTACCATCGCAACCCCGCTGGGA 1294

Db 599 AGTATGGATGGATGAATTTAAAGAGCTTACTACCATCGCAACCCCGCTGGGA 658

QY 1295 ACCTTTTGGGGATGTGACAGAGAGGAGCAGCTCCGGGCAAGCTCCAGTGTAAAGACTT 1354

Db 659 ACCTTTTGGGGATGTGACAGAGAGGAGCAGCTCCGGGCAAGCTCCAGTGTAAAGACTT 718

QY 1355 CAAGTGGTCTTGGAGACTGTGTATCCAGAACTGTCATGTCCTGAGGACAGGCTGGCTT 1414

Db 719 CAAGTGGTCTTGGAGACTGTGTATCCAGAACTGTCATGTCCTGAGGACAGGCTGGCTT 778

QY 1415 CTTTCGGGATGCTCCAGAAACAAAGACTTAACAGACTACTGCTTTGACTATAACCTCCCGA 1474

Db 779 CTTTCGGGATGCTCCAGAAACAAAGACTTAACAGACTACTGCTTTGACTATAACCTCCCGA 838

QY 1475 TGAACACCAAGTGTGGGACACAGGTCATTTGTACTCTGTCTGTCAGGATGGGCCAGAA 1534

Db 839 TGAACACCAAGTGTGGGACACAGGTCATTTGTACTCTGTCTGTCAGGATGGGCCAGAA 898

QY 1535 TCAGTCTTTCAGTACACGTCCTCCAGAAAGAAATACGCTATAACACCCAGGCTGAGGG 1594

Db 899 TCAGTCTTTCAGTACACGTCCTCCAGAAAGAAATACGCTATAACACCCAGGCTGAGGG 958

QY 1595 CTGCTTGTCTGGAAGCAGGAATGATACCTTATCATGCTCTCTGCGAAGAACTGC 1654

Db 959 CTGCTTGTCTGGAAGCAGGAATGATACCTTATCATGCTCTCTGCGAAGAACTGC 1018

QY 1655 CCCAGAGAATCAGAAGTTTCTTTCGAGGAGGATGGATCTTTATTTCCAGAACTGCTCA 1714

Db 1019 CCCAGAGAATCAGAAGTTTCTTTCGAGGAGGATGGATCTTTATTTCCAGAACTGCTCA 1078

QY 1715 GAAATGTCTCAGGCTCGAGGAGGAGTGCAGTGCAGTGTTCGTTCCACTCTTACGAGA 1774

Db 1079 GAAATGTCTCAGGCTCGAGGAGGAGTGCAGTGCAGTGTTCGTTCCACTCTTACGAGA 1138

QY 1775 CTCACCAACTCGGATCATCAGAAATGTTCTTCAAAGAGCGCATGTTATGAAGCCTCGT 1834

Db 1139 CTCACCAACTCGGATCATCAGAAATGTTCTTCAAAGAGCGCATGTTATGAAGCCTCGT 1198

QY 1835 GTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAG 1894

Db 1199 GTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAG 1258

QY 1895 CTAGCAGTACACGACACCCACCAAAACCTAGGCTGCTTGGAGGCAATCATT 1954

Db 1259 CTAGCAGTACACGACACCCACCAAAACCTAGGCTGCTTGGAGGCAATCATT 1318

QY 1955 TTGCCATTTGTGAAGTTGTGTTGGATTAGTAAATAATGCAATAGCTTTGTACTTATT 2014

Db 1319 TTGCCATTTGTGAAGTTGTGTTGGATTAGTAAATAATGCAATAGCTTTGTACTTATT 1378

QY 2015 TTGAGAACTTTTAAATGTTCCAAATAACCTATTTTCAAAGGGTAATCGTAAGATGTTA 2074

Db 2074 TTGAGAACTTTTAAATGTTCCAAATAACCTATTTTCAAAGGGTAATCGTAAGATGTTA 2074

Db 1379 TTGAGAACTTTTAAATGTTCCAAATACCCCTATTTTCAAGGGTAATCGTAAGATGTTA 1438
Qy 2075 ACCCTTGGTATTAGAAAAATTAACCTTATTAATATTTTCTATCAARAWRWA 2129
Db 1439 ACCCTTGGTATTAGAAAAATTAACCTTATTAATATTTTCTATCAARAWRWA 1493

RESULT 10

AA78414
ID AA78414 standard; cDNA; 1517 BP.

XX AC AA78414;

XX DT 20-NOV-2000 (first entry)

XX DE Human secreted protein gene 34 SEQ ID NO:44.

XX KW Human; secreted protein; cytostatic; antianaemic; antidiabetic;
KW antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
KW antiporiatic; antiangiogenic; cardiac; anti-HIV; nootropic;
KW neuroprotective; antimicrobial; antiparkinsonian; cancer;
KW immune system disorder; angiogenesis; hyperproliferative disorder;
KW cardiovascular disorder; apoptosis; neurological disease;
KW infectious disease; wound healing; ss.

XX OS Homo sapiens.

XX PN WO200035937-A1.

XX XX 22-JUN-2000.

XX PF 16-DEC-1999; 99WO-US29950.

XX PR 17-DEC-1998; 98US-0112809.

XX PR 18-DEC-1998; 98US-0113006.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Ebner R, Rosen CH, Endress GA, Soppet DR, Ni J;
PI Duan DR, Moore PA, Shi Y, Latleur DW, Olsen HS, Florence K;

XX PI WPI; 2000-431566/37.

XX DR P-PSDB; AAB24470.

XX PT Forty seven human nucleic acids encoding secreted proteins, useful in
PT the treatment, prevention and diagnosis of cancers, disorders of the
PT immune system, angiogenesis disorders, neurological diseases and
PT hyperproliferative disorders -

XX PS Claim 1; Page 465-466; 562pp; English.

XX CC The polynucleotide sequence given in AAA78381 to AAA78432 encode the
CC human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: cytostatic; antianaemic;
CC antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC antiporiatic; antipsoriatic; antiangiogenic; cardiac; anti-HIV;
CC nootropic; neuroprotective; antimicrobial and antiparkinsonian.
CC Human secreted protein polynucleotides, polypeptides, antagonists and/or
CC agonists may be useful in treating, preventing, and/or diagnosing other
CC diseases, disorders, and/or conditions such as: (a) cancers; (b)
CC disorders of the immune system; (c) angiogenesis disorders; (d)
CC hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
CC associated with increase apoptosis; (g) neurological diseases; and
CC (h) infectious diseases. They are also used to promote wound healing.
CC AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC exemplification of the present invention.

XX SQ Sequence 1517 BP; 366 A; 360 C; 363 G; 424 T; 4 other;

Query Match

Best Local Similarity 35.58; Score 1011.8; DB 21; Length 1517;

Matches 1007; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

RESULT 11

AA568954

ID AAS68954 standard; cDNA; 1069 BP.

Qy 1115 GATCTGGCAGTGTGGTGGGTTCTGAAACACACCCATGTTCCATGTTGGCCATGTTT 1174
Db 479 GATCTGGCAGTGTGGTGGGTTCTGAAACACACCCATGTTCCATGTTGGCCATGTTT 538
Qy 1175 CCCAAGCAGCTCCCTACTCCCGCACAGGCTCTGGCCAAAGTGTTCGTGCGCTGA 1234
Db 539 CCCAAGCAGCTCCCTACTCCCGCACAGGCTCTGGCCAAAGTGTTCGTGCGCTGA 598
Qy 1235 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGA 1294
Db 599 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGA 658
Qy 1295 ACCTTTTGGGATGTGACAGAGAGGAGCAGCTCCGGACAAAGCTCCAGTGTAAAGACTT 1354
Db 659 ACCTTTTGGGATGTGACAGAGAGGAGCAGCTCCGGACAAAGCTCCAGTGTAAAGACTT 718
Qy 1355 CAAGTGGTCTTGGAGACTGTGTATCCAGAACTGTCATGTCCTGAGGACAGGCTTGGCTT 1414
Db 719 CAAGTGGTCTTGGAGACTGTGTATCCAGAACTGTCATGTCCTGAGGACAGGCTTGGCTT 778
Qy 1415 CTTTGGGATGTCTCCAGAAACAAAGACTTAACAGACTACTGCTTTGACTATTAACCTCCCGA 1474
Db 779 CTTTGGGATGTCTCCAGAAACAAAGACTTAACAGACTACTGCTTTGACTATTAACCTCCCGA 838
Qy 1475 TGAACACAGATTTGGGACACAGGTCATTTCTGTACCTCTGTCTGATGGGATGGGCCAGAA 1534
Db 839 TGAACACAGATTTGGGACACAGGTCATTTCTGTACCTCTGTCTGATGGGATGGGCCAGAA 898
Qy 1535 TCAGTTTTTGGAGTACAGCTCCAGAAAGAAATACGTATAACACCCACAGCTGAGGG 1594
Db 899 TCAGTTTTTGGAGTACAGCTCCAGAAAGAAATACGTATAACACCCACAGCTGAGGG 958
Qy 1595 CTGCATTGCTGTGAAAGCAGGAATGATACCTTTATCATCATCTCTGCGAAGAACTGC 1654
Db 959 CTGCATTGCTGTGAAAGCAGGAATGATACCTTTATCATCATCTCTGCGAAGAACTGC 1018
Qy 1655 CCAGAGAAATCAGAACTTCACTTTGCGAGGAGTGGATCTTTATTTACGAAACAGTCCAA 1714
Db 1019 CCAGAGAAATCAGAACTTCACTTTGCGAGGAGTGGATCTTTATTTACGAAACAGTCCAA 1078
Qy 1715 GAAATGCTCCAGGCTGCGAGGAGAGTGCAGTGTTCCTTCCACTCTTACAGAA 1774
Db 1079 GAAATGCTCCAGGCTGCGAGGAGAGTGCAGTGTTCCTTCCACTCTTACAGAA 1138
Qy 1775 CTCACCAACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATGTTATGAAGCCCTCGT 1834
Db 1139 CTCACCAACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATGTTATGAAGCCCTCGT 1198
Qy 1835 GTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAG 1894
Db 1199 GTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAAAGACTTAG 1258
Qy 1895 CTAAGCAGTGCACAGAACCCACCAAACTAGGCTGCATTTGTAAGAGGCAATCAAT 1954
Db 1259 CTAAGCAGTGCACAGAACCCACCAAACTAGGCTGCATTTGTAAGAGGCAATCAAT 1318
Qy 1955 TTGCCAATTTGTAAGAGTGTGTTGGATTTAGTAAATGTAATAAGCTTTGTACTTATT 2014
Db 1319 TTGCCAATTTGTAAGAGTGTGTTGGATTTAGTAAATGTAATAAGCTTTGTACTTATT 1378
Qy 2015 TTGAGAACTTTTAAATGTTTCAAAATACCCCTATTTTCAAGAGGTAATCGTAAGATGTTA 2074
Db 1379 TTGAGAACTTTTAAATGTTTCAAAATACCCCTATTTTCAAGAGGTAATCGTAAGATGTTA 1438
Qy 2075 ACCCTTGGTATTAGAAAAATTAACCTTATTAATATTTTCTATCAARAWRWA 2129
Db 1439 ACCCTTGGTATTAGAAAAATTAACCTTATTAATATTTTCTATCAARAWRWA 1493

Human; breast cancer; cell marker; cytostatic; ss.
Homo sapiens.
WO200151628-A2.
19-JUL-2001.
10-JAN-2001; 2001WO-US00798.
14-JAN-2000; 2000US-0176077.
14-MAR-2000; 2000US-0189167.
24-MAR-2000; 2000US-0192099.
29-MAR-2000; 2000US-0193480.
15-MAY-2000; 2000US-0205210.
09-JUN-2000; 2000US-0211315.
25-JUL-2000; 2000US-0220534.
(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
Lillie J, Xu Y, Wang Y, Steinmann K;
WPI; 2001-451856/48.
New peptide useful as a marker for the diagnosis of breast cancer -
Claim 1; Page 3592-3593; 3695pp; English.
The invention relates to human breast cancer expressed polynucleotides
(AAL07544-AAL26789) and methods of assessing whether a patient is
afflicted with breast cancer by examining the correlation between the
expression of certain markers and the cancerous state of breast cells.
The polynucleotides and encoded polypeptides are potential markers for
detecting, diagnosing, monitoring, characterizing treating and
potentially preventing breast cancer. The polynucleotides and encoded
polypeptides are also useful for isolating compounds with cytostatic
activity.
Sequence 883 BP; 273 A; 172 C; 134 G; 289 T; 15 other;
Query Match 26.5%; Score 756.2; DB 22; Length 883;
Best Local Similarity 98.6%; Pred. No. 3.8e-128;
Matches 788; Conservative 5; Mismatches 3; Indels 3; Gaps 3;
QY 2031 TGTTCACAAATACCTATTTTCAAGGGTAATCGTAAGATGTTAAACCCCTTGGTATTAGA 2090
DB 873 TGTTCACAAATACCTATTTTCAAGGGTAATCGTAAGATGTTAAACCC-TGGTATTAGA 815
QY 2091 AAATTAACCTTATATATTTTCTATCAARAWAWATTTTACAGTCGTGCTTTTAC 2150
DB 814 AAATTAACCTTATATATTTTCTATCAAGATGTTATTTACAGTCGTGCTTTTAC 755
QY 2151 TCTCATTAGCAAAAAGATTAAGATTTTATTTGGTATTACAGGATCCAGGTACGA 2210
DB 754 TCTCATTAGCAAAAAGATTAAGATTTTATTTGGTATTACAGGATCCAGGTACGA 696
QY 2211 AGATATCTGCATGGTGGGAATCAGGTTCAAGCAACCTACTTTGCATTAACGTATATAC 2270
DB 695 AGATATCTGCATGGTGGGAATCAGGTTCAAGCAACCTACTTTGCATTAACGTATATAC 637
QY 2271 CTGAGTCGGGGTTAAAGTTTCCAGATAGAGAGACTGTCTAGGAAACATTTGATT 2330
DB 636 CTGAGTCGGGGTTAAAGTTTCCAGATAGAGAGACTGTCTAGGAAACATTTGATT 577
QY 2331 GATTATTACAGTCATGAGATCTTCTAGATGTTATTTTAAAGATGCTTTTGGTTAT 2390
DB 576 GATTATTACAGTCATGAGATCTTCTAGATGTTATTTTAAAGATGCTTTTGGTTAT 517
QY 2391 GTGTGTGTACACAGTTTAAACCTCCATAATGTTTCATGTACAGCAAGAGGACTAACCAAA 2450
DB 516 GTGTGTGTACACAGTTTAAACCTCCATAATGTTTCATGTACAGCAAGAGGACTAACCAAA 457
QY 2451 GCTGAAATCTCAGAGAAACAATTTGCTTTTACTTAAGCTGAGTCAACTTGGAGCGCACTTCT 2510

DB 456 GCTGAAATCTCAGAGAAACAATTTGCTTTTACTTAAGCTGAGTCAACTTGGAGCGCACTTCT 397
QY 2511 AACAAATGCCGCACTGTAGTGTGCTGTCTTACCACATATGACTTTTAAACATGTTTATAT 2570
DB 396 AACAAATGCCGCACTGTAGTGTGCTGTCTTACCACATATGACTTTTAAACATGTTTATAT 337
QY 2571 CATTTTAAATTTTATGATACGCTAGTGTGTCAGGAGAAATGTAATGTTCTATATGAAAT 2630
DB 336 CATTTTAAATTTTATGATACGCTAGTGTGTCAGGAGAAATGTAATGTTCTATATGAAAT 277
QY 2631 CCTTTTCAAGTTTGTTCATTAATAACAGTATTATTTTAAACAGCGTTAGAGTTGTG 2690
DB 276 CCTTTTCAAGTTTGTTCATTAATAACAGTATTATTTTAAACAGCGTTAGAGTTGTG 217
QY 2691 CTGCTCAACTGCTGTGAAATTTCTCTGAGTAAATCTGATTTGTGAATGATCCAGACC 2750
DB 216 CTGCTCAACTGCTGTGAAATTTCTCTGAGTAAATCTGATTTGTGAATGATCCAGACC 157
QY 2751 AACCTGAGATTTTGTCAACCTGATTAAAGTCAATATGAATGATTAAAAAGATGTGAGAAC 2810
DB 156 AACCTGAGATTTTGTCAACCTGATTAAAGTCAATATGAATGATTAAAAAGATGTGAGAAC 97
QY 2811 AAAAAAAAAAAAAAAAAAAAAA 2829
DB 96 ACTGAAAAAAAAAAAAAAAAAAAA 78
RESULT 13
AAA02624
ID AAA02624 standard; cDNA; 592 BP.
XX AAA02624;
XX AC
DT 19-MAY-2000 (first entry)
XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:2615.
DE Human; colon cancer; tumour; diagnosis; gene expression product;
XX probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
OS Homo sapiens.
XX WO958675-A2.
PN 18-NOV-1999.
PD 13-MAY-1999; 99WO-US10602.
XX 14-MAY-1998; 98US-0085426.
PF 15-MAY-1998; 98US-0085537.
XX 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Cckanjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX WPI; 2000-126369/11.
DR Polynucleotide library used to determine cancerous states of mammalian
XX cells -
PT Claim 1; Page 1061-1062; 1097pp; English.
PS AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
XX


```
Db 484 CCAACAAGACTTAGCTAAGCAGTGACAGAACCCACCAAAAAAATAGGCTTGCAATTGCTT 543
Qy 1939 TGAAGAG 1945
Db 544 TGAAGAG 550
RESULT 15
AAS24233
ID AAS24233 standard; cDNA; 544 BP.
XX AC AAS24233;
XX
XX 07-NOV-2001 (first entry)
DE Human ovarian PCR-subtracted cDNA library clone #414.
XX
XX Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss;
KW gene therapy; cytostatic; T-cell expansion; nucleic acid hybridisation;
KW primer; probe.
XX
XX Homo sapiens.
OS
XX
XX WO200157207-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 05-FEB-2001; 2001WO-US03733.
XX
XX 04-FEB-2000; 2000US-0180403.
PR
XX 28-MAR-2000; 2000US-0192745.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Algate PA, Mannion J;
PI
XX
XX WPI; 2001-488879/53.
DR
XX
XX New polynucleotides encoding ovarian tumour proteins, useful for
PT treating ovarian cancer, and as probes, primers, and markers of cancer
PT progression.
XX
XX Example 1; page 179; 378pp; English.
XX
XX The invention comprises compositions used for the therapy and diagnosis
CC of ovarian cancer. The compositions comprise one or more ovarian tumour
CC proteins, their associated polynucleotides, or immunogenic portions of
CC the proteins. The ovarian tumour polynucleotides and polypeptides are
CC useful for stimulating and/or expanding T cells specific for a tumour
CC protein. They are also useful for inhibiting the development of cancer in
CC a patient with an ovarian tumour DNA or protein by incubating isolated
CC T-cells allowing them to proliferate, and administering to the patient.
CC The sequences can be used as markers for cancer, for example, to monitor
CC ovarian cancer progression. Probes and primers are useful in nucleic acid
CC hybridisation, in detecting the presence of complementary sequences in a
CC given sample, for preparing mutant species and for preparing other
CC genetic constructions. Sequences AAS23820-AAS25231 and AAS25328-AAS25549
CC represent human ovarian tumour protein cDNA clones.
XX
XX Sequence 544 BP; 159 A; 135 C; 130 G; 119 T; 1 other;
SQ
Query Match 18.5%; Score 526.4; DB 22; Length 544;
Best Local Similarity 99.6%; Pred. NO. 2e-86;
Matches 538; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 1407 CCTGGCTTCTCGGGATGCTCCAGAACAAAGAGCTAACGACTACTGCTTTGACTATAAC 1466
Db 1 CCTGGCTTCTCGGGATGCTCCAGAACAAAGAGCTAACGACTACTGCTTTGACTATAAC 60
Qy 1467 CCTCCCGATGAACACAGATTGTGGGACACCCAGGTCAATCTGTACCTCTGTATGGGATG 1526
Db 61 CCTCCCGATGAACACAGATTGTGGGACACCCAGGTCAATCTGTACCTCTGTATGGGATG 120
```

```
Qy 1527 GGCCAGAAATCAGTTTTTTCGAGTACACGTCCCGAGAAAGAAATACGCTATTAACACCCACCAG 1586
Db 121 GGCCAGAAATCAGTTTTTTCGAGTACACGTCCCGAGAAAGAAATACGCTATTAACACCCACCAG 180
Qy 1587 CCTGAGGGCTGCATTGCTGTGGAGCAGGAATGGATACCCCTTATCATGCATCTCTGGGAA 1646
Db 181 CCTGAGGGCTGCATTGCTGTGGAGCAGGAATGGATACCCCTTATCATGCATCTCTGGGAA 240
Qy 1647 GAAACTGCCCCAGAGAAATCAGAAAGTTTCATCTTGCAGGAGGATGATCTTTTATTTTCAGAA 1706
Db 241 GAAACTGCCCCAGAGAAATCAGAAAGTTTCATCTTGCAGGAGGATGATCTTTTATTTTCAGAA 300
Qy 1707 CAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAAGGAGTCGAGTGACACAGTTTTCGTTCCACTC 1766
Db 301 CAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAAGGAGTCGAGTGACACAGTTTTCGTTCCACTC 360
Qy 1767 TTACGAGACTGCACCAACTCCGATCATCAGAAATGGTTCTTCAAGAGCGCATGTTATGA 1826
Db 361 TTACGAGACTGCACCAACTCCGATCATCAGAAATGGTTCTTCAAGAGCGCATGTTATGA 420
Qy 1827 AGCCTCGTGTATCAAGGAGGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAA 1886
Db 421 AGCCTCGTGTATCAAGGAGGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACAA 480
Qy 1887 AGACTTAGCTAAGCAGTACACAGAACCCACCAAAAACTAGGC-TGCATTGCTTTGAAGAG 1945
Db 481 AGACTTAGCTAAGCAGTACACAGAACCCACCAAAAACTAGGCTTGCATTGCTTTGAAGAG 540
```

Search completed: December 7, 2003, 13:57:46
Job time : 742.064 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 13:31:08 ; Search time 161.847 Seconds
(without alignments)
7772.402 Million cell updates/sec

Title: US-10-074-527-1
Perfect score: 2850
Sequence: 1 ccggctcggtaccactataa.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0.

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	496.2	17.4	1737	4	US-09-217-306B-1
2	285.6	10.0	1920	3	US-09-376-856-1
3	284	10.0	3332	3	US-09-347-488-1
4	253.8	8.9	1521	4	US-09-795-926-30
5	253.8	8.9	1812	4	US-09-795-926-42
6	253.8	8.9	3896	4	US-09-795-926-44
7	232.2	8.1	1617	2	US-08-967-508-18
8	232.2	8.1	1617	3	US-08-967-506-18
9	232.2	8.1	1617	5	PCT-US94-02552-18
10	232.2	8.1	1680	2	US-08-967-508-11
11	232.2	8.1	1680	3	US-08-967-506-11
12	232.2	8.1	1680	5	PCT-US94-02552-11
13	232.2	8.1	2294	2	US-08-967-508-10
14	232.2	8.1	2294	3	US-08-967-506-10
15	232.2	8.1	2294	5	PCT-US94-02552-10
16	222	7.8	3889	2	US-08-648-298-1
17	201.4	7.1	810	4	US-09-795-926-26
18	198.8	7.0	1608	4	US-09-795-926-38
19	198.8	7.0	1608	4	US-09-795-926-28
20	198.8	7.0	1899	4	US-08-785-926-40
21	134.8	4.7	675	4	US-09-795-926-22
22	134.8	4.7	966	4	US-09-795-926-34
23	97.6	3.4	903	4	US-09-620-312D-938
24	87.4	3.1	322	1	US-08-594-031-61
25	76.4	2.7	262	1	US-08-594-031-62
26	74.2	2.6	2416	4	US-09-620-312D-112
27	69.2	2.4	339	4	US-09-795-926-24

ALIGNMENTS

RESULT 1

US-09-217-306B-1
; Sequence 1, Application US/09217306B
; Patent No. 6465220
; GENERAL INFORMATION:
; APPLICANT: Hassan, Helle
; APPLICANT: Clausen, Henrik
; APPLICANT: Bennett, Eric P.
; TITLE OF INVENTION: Glycosylation Using GalNac-T4 Transferase
; FILE REFERENCE: 8850*1
; CURRENT APPLICATION NUMBER: US/09/217,306B
; CURRENT FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-217-306B-1

Query Match	17.4%	Score 496.2;	DB 4;	Length 1737;
Best Local Similarity	62.3%	Pred. No. 5.2e-96;		
Matches	796;	Conservative	0;	Mismatches 478;
				Indels 3;
				Gaps 1;
Qy	300	GCCTGGCGCGCGGGCGAGCGGTGCGGCTGCAGCTGCAGGGCGAGGAGCTGCGGCTG	359	
Db	217	GCACCTGGGGAGTGGGGAAAGCCAGCAAACTCCAGCTCAACGAGGATGAATGAAGCAG	276	
Qy	360	CAGGAGGAGCGTGGCGGTGCACACAGATTAACATCTACCTCAGCGACCCGCTCACTG	419	
Db	277	CAAGAAGAACTCAATTGAGAGATACGCCATCAATATTACCTCAGTACAGGATTTCCCTG	336	
Qy	420	CACCGCGCTGCGCGGCGCTGGAACCGCTGCAAGAGAGAAATATGATATGAT	479	
Db	337	CATCGACATAGAGGATAAAGATGATAGTGAAGTCCCAAGAGTTCAACTATAGG	396	
Qy	480	AATTTGCCAGGACATCTGTTATCATAGCATTTTATAAGAGCTGGTCAACTCTCTT	539	
Db	397	ACACTTCTTACCACCTCTGTTATCAITGTTCTTATAACGAGAGCTGGTCACTTCTCTC	456	
Qy	540	CGGACAGTTTACAGTCTCTTGGAGATCCCGGATATCTCGTAGAGAGTATCTT	599	
Db	457	CGTACCATTACAGTGTTTTAGAAACTTCTCTCGAGTCTTTTGAAGAGATCATCTT	516	
Qy	600	GTAGATGACTACAGTATAGAGAGACCTGGAAGAGCGCTTGGCCAAATGAGCTTCCGGA	659	
Db	517	GTGATGACTTGTAGTGACAGAGTTTATTGAAGACACAACTTGAACATTCAGCAAT	576	
Qy	660	CTGCCAAGGTGCGCTGTATCCGCGCAACAGAGAGAGCGCTGTGTCGAGCCGCTG	719	

QY 1290 TTGGAACCTTTTGGGGATGTGACAGAGAGG-----AGCAGCTCCGGGACAAAGCTC 1340
DB 1381 TCCTTGAGCAAGGCTGAGAGCCAGAGCTGCTATGGAACGCTTGAGCTGCAAGAGAGACTG 1440
QY 1341 CAGTGTAAAGACTTCAAGTGGTCTTCTGGAGACTGTGTATCCAGAACTGCATGTGCCCTGAG 1400
DB 1441 GGTGTGGGACATTCACCTGGTCTTCTGGCTAAATGTCTACCTTGAGCTGTACCCATCTGAA 1500
QY 1401 GACAGGCTGGCTCTTTCGGAGTCTCCAGAAACAAAGGACTAACAGACTACTGTCTTTGAC 1460
DB 1501 CCCAGGCCAGTTCTCTGGAAGCTCCACAACACTGGACT-----TGGGCTCTGT 1551
QY 1461 TATAACCTCCGATGAAACACAGATTGTGGGACACAGGTCATCTGTACCTCTGTTCAT 1520
DB 1552 GCAGACTGCCAGGAGAGGGGACATCTGGGCTGTCCCATGTGTGTGGCTCTCTTGCAGT 1611
QY 1521 GGGATGGGCCAGAAATCAGTTTTCAGGTACAGCTCCAGAAAGAAATACGCTATAACACC 1580
DB 1612 GACAGCCGGCAGCACAAGTACCTGCAGACACACAGCAAGGAGATTCATCTTGGCAGC 1671
QY 1581 CACCAGCTGAGGCTGCTATGTGTGGAAGCAGGAATGGATACCTTTATCATGCACTC 1640
DB 1672 CCACAGC---ACCTGTGCTTGTCTCAGCAGGAGCAGGTGATCTTTCAGAACTGCACG 1728
QY 1641 TCGAAGAACTGCCCCAGAGAAATCAGAACTTCAATCTTTCAGGAGGATGGATCTTTATTT 1700
DB 1729 GAGGAAGGCTTGCCCATCCACAGCAGCACTGGGACTTCCAGGAGATGGGATGTGTC 1788
QY 1701 CAGGAAGCTTCAAGAAATCTGCCAGCTGCCAGGAGGAGTCGAGTGACAGTTT 1756
DB 1789 CACATCTTTCTGGGAATGCATGGAAGCTGTGTGCAAGAAACAAATNAAGATTT 1844

RESULT 3

US-09-347-488-1
; Sequence 1, Application US/09347488
; Patent No. 6239266
; GENERAL INFORMATION:
; APPLICANT: Pribill, Ingrid
; APPLICANT: Munroe, David
; TITLE OF INVENTION: ZAP-3 TUMOR ASSOCIATED GENES AND THEIR
; FILE OF INVENTION: USES
; FILE REFERENCE: GEN-2PRV
; CURRENT APPLICATION NUMBER: US/09/347,488
; CURRENT FILING DATE: 1999-07-02
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3332
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (368)...(2284)
US-09-347-488-1

Query Match 10.0%; Score 284; DB 3; Length 3332;
Best Local Similarity 53.8%; Pred. No. 7.4e-51;
Matches 740; Conservative 0; Mismatches 600; Indels 36; Gaps 6;
QY 399 CTCAGCGACCGCATCTCACTGCAACCGCGCTCCCGAGCGCTGGAAACCGCTGTGCMAA 458
DB 854 CTCAGTCCCGCATCCCTCCAGAGGGCTCTCCCGAGGTGCGGCACCCACTGTG---T 910
QY 459 GAGAAGAAATATGATATGATTAATTTGCCAGAGACATCTGTTATCATAGCATTTTATAT 518
DB 911 CTGACGACGACCCCTCAGGACAGCCCTGCCACAGCCAGCGCTATCTCTGTTTCCATGAT 970
QY 519 GAAGCTGTGTCACCTCTCCCTTCGGACAGATTTACAGTGTCTCTCAGACATCCCGGATATC 578
DB 971 GAGGCTGTGTCACCTCTCCCTGGGACTGTACAGCATCTCTCGACACAGTGTGCCAGGGCC 1030

QY 579 CTGCTAGAAAGAGTGTATCTCTTGTAGATGACTACAGTGTATAGAGAGCACTTGAAGGAGCGC 638
DB 1031 TTCTCTGAAGGAGATCATCTCTCGTGGACGACTCAGCCAGCAAGCAACTCAAGCTCTGCT 1090
QY 639 TTGGCCAAATGAGCTTTTCGGGACTGCCCAAGTGGCTGATCCGCGCAACAGAGAGAG 698
DB 1091 CTGAGCGAATATGTGGCCAGGCTGGAGGGGTGAAGTTACTCAGGAGCAACAAGAGGCTG 1150
QY 699 GGCTCTGTGCGAGCCCGGCTGTCTGGGGGCGCTCTGCGCGAGGGGCGCATGTTCTTGACCTTC 758
DB 1151 AGTGCCATCAGGCGCCGATGTCTGGGGGCCACAGAGCCACCGGGATGTCTCGTCTTC 1210
QY 759 CTGCACTGTCTACGTGTAGTGCCACGAAGGGTGTGAGCGCTGTCTGCAAGAGATCCAT 818
DB 1211 ATGGATGCCCACTGCGAGTGCACCCAGGCTGGCTGGAGCCCTCTCTCAGCAGAATAGCT 1270
QY 819 GAAGAGAGTCCGAGTGTGTGCCGCTGATTTGATGTGATCTGACTGGAACACTCTCGAA 878
DB 1271 GGTGACAGAGCCGAGTGGTATCTCCGCTGATAGATGTGATGACTGGAAGACTTTCCAG 1330
QY 879 TACTGGGGAATCTCCGGGAGCCCGAGATCGCGGCTTTCGACTGGAGGCTGGTGTTCACG 938
DB 1331 TA---TTACCCCTCAAGGACCTGCGAGCTGGGGTGTGAGCTGGAAGCTGGATTTCCAC 1387
QY 939 TGGCACACAGTCTCTGAGAGGAGAGATACGGATGCAATCCCGCTGCGATGTCATCAGG 998
DB 1388 TGGGAACCTTTGCGAGAGCATGTGAGGAAGGCCCTCCAGTCCGCCATAAGCCCATCAGG 1447
QY 999 TCTCCAAACAATGGCTGGCTGGTGTCTGTGAGTAAGAAATATTTGAAATATCTGGGG 1058
DB 1448 AGCCCTGTGGTGGCCCGGAGAGGTGGTGGCCATGGACAGACATTACTTCCAAAACACTGGA 1507
QY 1059 TCTTATGATACAGGAATGGAAGTTTGGGAGAGGAAACCTCGAATTTTCTCTTAGGATC 1118
DB 1508 GCGTATGACTCTTATGTGCTGCGAGGTGGTGAAGAACTCGAATCTCTTCAAGGCC 1567
QY 1119 TGGCAGTGTGGTGGGTTCTGGAACAACACCCATGTTCCCATGTTGSCCATGTTTTCCTCC 1178
DB 1568 TGGCTCTGTGGTGGCTCTGTGAAATCTTCTCCCTGCTCTCGGTAGGACACATCTACCAA 1627
QY 1179 AGCAAGCTCCCTACTCTCCGCAACAGGCTCTGGCC-----ACAGTGTTCGTGCA 1229
DB 1628 AATCAGGATTCCTCATTTCCCTCCCTCGACAGGAGGCCACCTGAGGAACAGGGTTCGCA 1687
QY 1230 GCTGAAGTATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGCTGCCCGC 1289
DB 1688 GCTGAGACCTGCTGGGTCATTCAGAAACCTTCTACAGCATAGCCACAGAGGCTTC 1747
QY 1290 TTGGAACCTTTTGGGGATGTGACAGAGAGGA-----AGCAGCTCCGGGCAAGCTC 1340
DB 1748 TCCTTGAGCAAGGCTGAGAAGCCAGACTGCTGGAACGCTTGCAGCTGCAAGGAGACTG 1807
QY 1341 CAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCCCTGAG 1400
DB 1808 GGTGTGCGGACATTCACACTGGTCTTGGGCTAAATGTCTACCTGAGCTGTACCCATCTGAA 1867
QY 1401 GACAGGCTGGCTTCTTTCGGGATGCTCCAGAAACAAAGGACTAACAGACTACTCTGCTTTGAC 1460
DB 1868 CCCAGGCCAGATTTCTCTGGAAGCTCCACAACACTGGACT-----TGGGCTCTGT 1918
QY 1461 TATAACCTCCCGATGAAACCAAGATTGTGGGACACAGGTCATTTCTGTACCTCTGTTCAT 1520
DB 1919 GCAGACTGCCAGCAGAGGGGACATCTCTGGGCTGTCCCATGTTGGTGTGGCTCTTGCAGT 1978
QY 1521 GGGATGGGCGAGATCAGTTTTTCGAGTACACGCTCCAGAAAGAAATACGCTATAACACC 1580
DB 1979 GACAGCCGCGACCAACAGTACCTTGCAGACACACAGCAGGAAGGAGATTCATCTTGGCAGC 2038
QY 1581 CACAGCCTGAGGGCTGCAATGCTGTGGAAGCAGGAATGGATACCTCTATCATGCACTCTC 1640
DB 2039 CCACAGC---ACCTGTGCTTGTCTGTCAGCAGGACAGGTGATTTCTTCAGAACTGCAGC 2095
QY 1641 TGCGAAGAAACTGCCCCCAGGAATCAGAAATTCATCTTGAGGAGGATGGATCTTTATTT 1700

```
Db 2096 GAGGAGGCGCTGCCATCCACAGCAGCACTGGGACTTCACAGAGAAATGGGATGTC 2155
Qy 1701 CACGAACAGTCCAGAAATGTGTCAGGCTGCGAGGAGAGATCGAGTGACAGTTT 1756
Db 2156 CACATTTCTTCTGGAAATGCATGGAAGCTGTGTGCAAGAAACAAATGAAGATTT 2211

RESULT 4
US-09-795-926-30
; Sequence 30, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME
; CURRENT FILING DATE: 2001-02-28
; PRIOR FILING DATE: 2000-02-29
; PRIOR FILING DATE: 2000-03-02
; PRIOR FILING DATE: 2000-03-02
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PASTSEQ for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 1521
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-795-926-30

Query Match 8.9%; Score 253.8; DB 4; Length 1521;
Best Local Similarity 55.5%; Pred. No. 1.3e-44;
Matches 561; Conservative 0; Mismatches 432; Indels 18; Gaps 3;

Qy 388 TTAACATCTACCTCAGCGACCGCATCTCACTGCACCGCGCTGCCCGAGCGCTGGAACC 447
Db 47 TTAACATCTACCTCAGCGATTAATAATCTCTTGAATCGCTCTCTCCAGATATCCGGACC 106
Qy 448 CGCTGTGCAAGAGAGAAATATGATTATGATAATTTGCCAGGACATCTGTATATAG 507
Db 107 CAACTGCAACAGCAGCGCTACCT---GGAGACACTTCCCAACACAGCATCATATCC 163
Qy 508 CATTTTATATGAAGCTCGTCAACTCTCTCTCGGACAGTTTACAGTGTCTTGAACAT 567
Db 164 CTTTCCCAACAGAGGCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 223
Qy 568 CCCCAGATATCTCTAGAGAGATGATCTCTGTAGATGACTACAGTATAGAGACACC 627
Db 224 CGCTCCAGAGCTGGTCCCGAGATGTACTGTGTCGACGACTTCAGTATCGAGAGACC 283
Qy 628 TGAAGAGCGCTTGGCCAAATGAGCTTTTGGGACTGCCCAAGGTGCGCTGATCCGCCCA 687
Db 284 TGAAGAGCGCTTGAAGAGCTACATGCGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 343
Qy 688 ACAAGAGAGAGGCTGTGTCAGCGCGCTGTCTGGGGCGCTCTGCGCGAGGCGCATG 747
Db 344 AGAAACGGAGAGGCTGTATAGGAGCCCGAATGCTGGGGCGCTCAGTGGCAACTCGGGATG 403
Qy 748 TTCTGACCTTCTCTGACTGTCACTGTGTAGTGCCACGAAGGGTGGCTGGAGCGCTGTGTC 807
```

```
Db 404 TCATCACATTTCTGGATTCCACTGTGAAGCAATGTCAATGGCTTCCCTCCCTTGTG 463
Qy 808 AGAGGATCCATGAAGAGGAGTGGCAGTGGTGGCCGGTGTGATGTGATGTGATGACTGGA 867
Db 464 ACGCATTTCTCGGAACCGCAAGACCAATTTGTGTGCGCGATGATGATGATTAATGACCATG 523
Qy 868 ACACCTTTCAATACCTCGGGAACTCCGGGAGAGCCCGCAGATCGGCGGTTCGACTGGAGGC 927
Db 524 ACGATTTTCGTACGAGACACAGGAGGGATGCCATCGGGGAGCCTTTGACTGGAGA 583
Qy 928 TGGTGTTCACGTGGCACACAGTTCTCTGAGAGGAGAGATACGATGCAATCCCGCTCG 987
Db 584 TGTACTACAAGCGGATCCCGATCCCT-----CCAGAACTGCAGAAAGCTGACCCAGCG 637
Qy 988 ATGTCAATCAGTCTCCAAACATGGCTGTGGGCTGTGCTGTGATGAGAAATATTTT 1047
Db 638 ACCATTTGAGTCTCCCGTGTGATGGCGGTGGACTGTTCCGCGTGGATCGGAAGTGGTTCT 697
Qy 1048 AATATCTGGGCTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAGAAACCTCGAATTTT 1107
Db 698 GGAATCTGGCGGTATGACCCAGGCTTGGAGATCTGGGAGGGGAGCAGTATGAATCT 757
Qy 1108 CTTTAGGATCTGCAGTGTGTTGGGTTCTTGAAACACACCCATGTTCCCATGTTGGCC 1167
Db 758 CTTCAAGGTGTGATGTGTGGGGCGCATGAGGACATCCCTCTGCTCCAGGTGGGCC 817
Qy 1168 ATGTTTTCCTCCAGCAAGCTCCCTAC-----TCCCGCAACAGGCTCTGCCCAACA 1218
Db 818 ATATCTACAGGAATGATGTGCCCTCAAGGTCCCGCGGAGTCAAGCTGGCGCGGAACC 877
Qy 1219 GTCTTCGTGAGCTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1278
Db 878 TTAAGCGGTGGCGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 937
Qy 1279 CCGTGTCCCGCTTGGAACTTTTGGGATGTGACAGAGAGAGAGAGTCCGGGACAGC 1338
Db 938 CTGAATACCGCCACCTCTCCGCTGGGATGTGTCAGTCCAGAAAGCTCCGAGCTCCC 997
Qy 1339 TCCAGTGTAAAGACTTCAAGTGTCTTGGAGACTGTGATCCAGACTGC 1389
Db 998 TTAACCTGCAAGAGTTTCAAGTGGTTTATGACGAAGATAGCCTGGGACCTGC 1048
```

RESULT 5

```
US-09-795-926-42
; Sequence 42, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME
; CURRENT FILING DATE: 2001-02-28
; PRIOR FILING DATE: 2000-02-29
; PRIOR FILING DATE: 2000-02-29
; PRIOR FILING DATE: 2000-03-02
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
```

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 42

; LENGTH: 1812

; TYPE: DNA

; ORGANISM: homo sapiens

US-09-795-926-42

Query Match 8.9%; Score 253.8; DB 4; Length 1812;

Best Local Similarity 55.5%; Pred. No. 1.4e-44;

Mismatches 561; Conservative 0; Mismatches 432; Indels 18; Gaps 3;

```
Qy 388 TTAACATCTACCTCAGCGACCGCATCTCACTGCAACCGCGCTGCGGAGCGCTGGAAAC 447
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 338 TTAACATCTAGCTGATGATAAATCTCTTGAATCGCTCTCTCCAGATATCCGACCC 397
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 448 CGCTGTGCAAGAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAG 507
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 398 CAAACTGCAACAGCAAGCGCTACCT---GGAGACATCTCCCAACACAGCATCATCC 454
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 508 CATTTTATATGAAGCTGTGTAACCTCTCTTGGAGAGTTTACAGTGTCTTGAGACAT 567
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 455 CCTTCCAAACGAGGGTGGTCTCTCCCTCTCCGACCGCTCCACAGTGTGCTCAATCGCT 514
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 568 CCCGGATATCTGCTAGAGAGTATGATCTTGTAGATGACTACAGTATAGAGACCC 627
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 515 CGCTCCAGAGCTGTGCGGAGATTGATCTGTGCGAGACTTCAGTATCGAGAGACCC 574
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 628 TGAAGGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGGTGCGCTGATCCGCCCA 687
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 575 TGAAGAGCCTCTGAGAGACTACATGCGCTTTTCCCGAGTGTGAGGATCTTCGAACCA 634
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 688 ACAAGAGAGAGGCTCGTGGAGCGCGCTGTGCGGGCGTCTGCGGAGAGGGCGATG 747
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 635 AGAAACGGGAGGGCTGATAAGGACCGGAATGCTGGGGGCTCAGTGGCAACTGGGGATG 694
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 748 TTCTGACCTTCTGGAGTGTCACTGATGAGTCCACAGAGGGTGGCTGGAGCGCTGTGC 807
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 695 TCATCAATCTTGGATTCACTGTGAAGCAATGTCAACTGGCTTCCCGCTTGGTTG 754
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 808 AGAGGATCCATGAAGAGAGTGGCGAGTGTGTGCGCGGTGATTGATGATCGACTGGA 867
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 755 ACCGATCTCGGACCGCAAGACCAATGTGTGCGCGGATGATTGATTGATGACCATG 814
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 868 ACACCTTCGAATACCTGGGAACTCGGGAGCGCCAGATCGCGGTTTCAGCTGGAGGC 927
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 815 ACCACTTTCGATCGAGACACAGCAGCGGATGCCATGCGGGAGCCTTTGACTGGGAGA 874
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 928 TGGTGTTCAGTGCACACAGTTCCTGAGAGGAGAGATACGGATGCAATCCCGCTCG 987
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 875 TGACTACAAGCGGATCCCGATCCCT-----CCAGAACTGCAGAAAGCTGACCCAGCG 928
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 988 ATGTCTATCAGGTCTCCAAACAATGGCTGGTGGCTGTTTGTGTGAGTAAGAAATATTTG 1047
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 929 ACCATTTGAGTCTCCGCTGATGGCGGTGAGATGTTGCGCGTGGATCGGAAGTGTCT 988
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1048 AATATCTGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTT 1107
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 989 GGGAACTCGCGGGTATGACCCAGGCTTGAGATCTCGGGAGGGGAGCAGTATGAATCT 1048
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1108 CTTTAGGATCTGGCAGTGTGGGGTTCTGGAACACACCCATGTTCCCATGTTGGCC 1167
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1049 CTTCAAGGTGTGGATGTGGGGGCGCATGGAGGACATCCCGCTGCTCCAGGTTGGGCC 1108
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1168 ATGTTTTCCTCAAGAGCTCCCTAC-----TCCCGCAACAAGGCTCTGCCAACA 1218
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1109 ATATCTACAGAGATATGTCCTTACAGGTCCCGCGGAGTACGCTGGCGGGAAC 1168
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1219 GTGTTCTGAGCTGAAGTATGATGATGAATTTAAAGAGCTCTACTACCATCGCAAC 1278
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1169 TTAAGCGGGTGGCGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1228
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1279 CCGTGGCGCTTGGAAACCTTTTGGGGATGTGACAGAGAGAAAGAGCTCCCGGACAA 1338
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 1229 CTGAATACCCACCTCTCCGCTGGGGATGTCGAGTCCAGAAAAAGCTCCGAGCTCCC 1288
Qy 1339 TCCAGTGTAAAGACTTCAAGTGGTCTTCTTGAGACTGTGTATCCAGAACTGC 1389
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1289 TTAACGCAAGAGTTTCAAGTGGTATTATGACGAAGATAGCTGGGACCTGC 1339
```

RESULT 6

US-09-795-926-44

; Sequence 44, Application US/09795926

; Patent No. 6555669

; GENERAL INFORMATION:

; APPLICANT: Donoho, Gregory

; APPLICANT: Hilbun, Erin

; APPLICANT: Turner, C. Alexander Jr.

; APPLICANT: Friedrich, Glenn

; APPLICANT: Abuin, Alejandro

; APPLICANT: Zambowicz, Brian

; APPLICANT: Sands, Arthur T.

; APPLICANT: Walke, D. Wade

; APPLICANT: Wilganowski, Nathaniel L.

; APPLICANT: Hu, Yi

; APPLICANT: Kieke, James Alvin

; APPLICANT: Pottner, David George

; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

; FILE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME

; FILE REFERENCE: LEX-0144-USA

; CURRENT APPLICATION NUMBER: US/09795,926

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: US 60/185,920

; PRIOR FILING DATE: 2000-02-29

; PRIOR APPLICATION NUMBER: US 60/186,558

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: US 60/191,849

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 47

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 44

; LENGTH: 3896

; TYPE: DNA

; ORGANISM: homo sapiens

US-09-795-926-44

Query Match 8.9%; Score 253.8; DB 4; Length 3896;

Best Local Similarity 55.5%; Pred. No. 2e-44;

Mismatches 561; Conservative 0; Mismatches 432; Indels 18; Gaps 3;

```
Qy 388 TTAACATCTACCTCAGGACCGCATCTCACTGCAACCGCGCTGCGGAGCGCTGGAAAC 447
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 347 TTAACATCTACGTCAGTGTATAAATCTCTTGAATCGCTCTCTCCAGATATCCGACCC 406
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 448 CGCTGTGCAAGAGAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAG 507
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 407 CAAACTGCAACAGCAGCGCTACCT---GGAGACACTTCCCAACACAGCATCATATCC 463
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 508 CATTTTATATGAAGCTGTGTCACACTCTCTCTCGGACAGTTTACAGTGTCTTGGAGACAT 567
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 464 CTTTCCACACAGAGGCTGGTCTCTCTCTCTCGACCGTCCACAGTGTCTCAATCGCT 523
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 568 CCCCAGATATCTGTCTAGAGAAAGTGTATCTTGTAGATGATACATGATGATGAGAGACCC 627
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 524 CGCTCCAGAGCTGTGTCGCGAGATGTTACTGTGTCGACGACTTCAGTGTGAGAGACCC 583
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 628 TGAAGAGCGCTTGGCCCAATGAGCTTTCGGGACTGCGGAGTGGCGCTGATCCGCGCCA 687
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 584 TGAAGAGCGCTTGTGAAGACTACATGGCCCTTTTCCCAAGTGTGAGGATTCCTCGAACCA 643
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 688 ACAAGAGAGAGGCGCTGGTGGCGAGCCCGCTGTGCGGGCGCTCTGCGGAGGGCGATG 747
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 644 AGAAACGGGAGAGGCTGTATAAGGACCCGAATGCTGGGGGCTCAGTGGCAACTGGGGATG 703
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 748 TTCTGACCTTCTGGAGTGTCTCACTGTGAGTGCACAGAGGGTGGCTGGAGCCGCTGTGC 807
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



```
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-833-2210
; TELEFAX: 616-833-8897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-967-506-11

Query Match      8.1%; Score 232.2; DB 3; Length 1680;
Best Local Similarity 54.8%; Pred. No. 5.3e-40;
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;

Qy 472 ATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCCCTGGTCAA 531
Db 332 ATCCAGATAACTTCTACACCAAGTGTGGTGAATTTTCCACAATGAGGCTTGGAGCA 391

Qy 532 CTCTCTCCGACAGTTTACAGTGTCTTGGAGACATCCCGGATATCTGCTAGAGAAG 591
Db 392 CACTTCTCGCAACTGTCCATAGCGTCAATTAATCGCTCACCAGGACATGCTAGAGAAA 451

Qy 592 TGATCCTGTAGATGACTACAGTGTATGAGAGACCTGAAGAGCGCTTGGCCAAATGAGC 651
Db 452 TTGTTCTAGTAGATGATGCCAGTGAAGAGACTTTTAAAGAACCTCTAGAGAGTTACG 511

Qy 652 TTTCGGGACT---GCCCAAGGTGGCGCTGATCCCGGCCCAACAGAGAGAGCGGCTGGTGC 708
Db 512 TGAATAAATAAAGTACCCTTCACGTTCATCGAATGGAGCAGCGTCTGATGATCA 571

Qy 709 GAGCCCGCTCTCTGGGGCGTCTCGGAGGCGGATGTTCTGACCTTCTGACATGTC 768
Db 572 GAGCTAGTTAAAGGTGCTGCTGTCTAAAGGCCAAGTGATCACCTTTTACACGCGC 631

Qy 769 ACTGTGATGCCAGNAGGTGGCTGGAGCGCTGCTGCAGAGATCCATGAAGAGGT 828
Db 632 ACTGTGATGCCAGTGGGGTGGAGCGCTCTTTAGCCAGGATCAACATGACAGGA 691

Qy 829 CGGAGTGGTGTGCGCGGTGATGATGTGATGCTGGAACACCTTCGAATACCTGGGA 888
Db 692 AGACAGTGTCTGTCCCATCATAGATGTGATGATGATGATGATGATGATGATGATG 751

Qy 889 ACTCGGGAGGCCAGATCGCGGGTTCGATCGAGGCTGGTGTTCACGTGGGACACAG 948
Db 752 GTTCTGACATGACCTA---TGGCGGGTTCAACTGGAAGCTCAACTTTTCGCTGGTATCCTG 808

Qy 949 TTCTGTGAGGAGGAGG---ATACGGATGCAATCCCGCTGATGTCATCAGGTCTCAA 1005
Db 809 TTCCCCAAGAGAATGGACAGAGGAAGGTGATCGGACTCTTCCTGTGAGAACCTTA 868

Qy 1006 CAATGGCTGTGGGTGTTTCTGTGATGATGATGATGATGATGATGATGATGATGATG 1065
Db 869 CAATGGCAGGAGGCTTTTTCATAGACAGATTAATTTTTCAGAAATTTGACATATG 928

Qy 1066 ATACAGGAATGGAATTTGGGAGGAGAAAACCTCGAATTTCTTTAGGATCTGGCAGT 1125
Db 929 ATGCTGGAATGGATTTTGGGAGGAGAAAACCTAGAAAATTTCTTTAGGATTTGGCAGT 988

Qy 1126 GTGGTGGGTTCTGGNAACACACCCATGTTCCCATGTTGGCATGTTTTCCTCCCAAGCAAG 1185
Db 989 GTGGAGGAACCTTTGGAGATGTTTACTGCTCACATGTTGGACATGTTTTCGGAAGCTA 1048

Qy 1186 CTCCTACTCTCCGCAACAGGCTCTGGGCCAACAGTGT-----TCGTGCAG 1230
Db 1049 CACCTACAGCTTTCCAGGAGGACCGGGCAGATTAATCAATAAATAAACAAGCAGCTTG 1108

Qy 1231 CTGAAGTATGATGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGCCTGCT 1290
Db 1109 CAGAAGTATGATGATGAATTTCAAGAAATTTCTTCTATATAATTTCTCCAGGTGTACAA 1168

1291 TGGAACTTTTGGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAGCTCCAGTGTAAAG 1350
1169 AGGTAGATATGAGATATATCATCAAGACTTGGTCTAAGGCACAACTCCAATGCAGAC 1228
1351 ACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGTGATGCTGCTGAGGACAGGCTG 1410
1229 CATTCTCTGGTACCTAGAGATATTTATCCTGATTTCTCAGATTCTCTGTCACATTTCT 1288
1411 GCTTCTTGGGATGCTCCAGAACAAAGGACTACAGACTACTGCTTTGACTATAACCTC 1470
1289 CTTTGGGAGAGATAC-----GAAATGTGGAACAAATCAGTGTCTAGATAACATGGCTA 1342
1471 CCATGAAAACCAAGATTTGGGACACCAAGCTCATTTCTGTACCTCTGTCTGATGGGATGGCC 1530
1343 GAAAAGAGATGAAGAAGTTGGA-----ATTTTAACTGTCTGATGGTATGGAG 1390
1531 AGAATCAGTTTTTTCGAGTACACGTCCCAAGAAAGAAAT 1567
1391 GTAATCAGTTTTTCTCTTACACTGCACTGCAACCAAGAAAT 1427

RESULT 12
PCT-US94-02552-11
; Sequence 11, Application PC/TUS9402552
; GENERAL INFORMATION:
; APPLICANT: Elhammer, Ake P.
; APPLICANT: Homa, Fred L.
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNac:
; N-ACETYLGLYCOSYLTRANSFERASE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: The Upjohn Company, Corp. Intellectual
; ADDRESSEE: Property Law
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/02552
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-385-5210
; TELEFAX: 616-385-6897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; PCT-US94-02552-11

Query Match      8.1%; Score 232.2; DB 5; Length 1680;
Best Local Similarity 54.8%; Pred. No. 5.3e-40;
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;

Qy 472 ATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCCCTGGTCAA 531
Db 332 ATCCAGATAACTTCTACACCAAGTGTGGTGAATTTTCCACAATGAGGCTTGGAGCA 391
```

Qy	532	CTCTCCTTCGGACAGTTTACAGTGTCTTTGAGACATCCCCGGATATCTCTGTAGAGAAG	591
Db	392	CAC'TTCTCGCAACTGTCTCATAGCGTCA'TTAATCGCT'ACC'AAGGCA'CATGCTAGAGAAA	451
Qy	592	TGATCCTTTGATAGTACTACAGTGCATAGAGAGCACTCGAAGGCGCTTGCCTAATGAC	651
Db	452	TTGTTCTAGTAGATGATGCCAGTGAAGAGACTTTTAAAGAACCTCTTAGAGAGT'TACG	511
Qy	652	TTTCGGGACT---GCCCAAGGTGGCCCTGATCCCGGCCAAACAAGAGAGAGGGCCCTGGTGC	708
Db	512	TGAAAAAATTAAGTAGTACCCGGTTCACGTCATTCGAATGGAGCAGCTTCGTGATTGATCA	571
Qy	709	GAGCCCGCTCTCGGGGCGTCTCGGGGAGGGGGGATGTTCTGACCTTCTGTGACTGTCT	768
Db	572	GAGTAGGTTAAAAAGTGCTGCTGTCT'AAAGGCCAAGT'GATCACCTTTT'TAGACGCGC	631
Qy	769	ACTGTGAGTGCACGAAGGTGGCTGGAGCGCTGCTGCAGAGATCCATGAAGAGAGT	828
Db	632	ACTGTGAGTGCACAGTGGGTGGCTGGAGCGCTCTCTTAGCCAGGATCAACATGACAGGA	692
Qy	829	CGGCAGTGTGTGCCCCGGTGATTGATGTGATCGACTGGAAACACCTTCGAATACCTCGGGA	888
Db	692	AGACAGTGTGTCTCCCATCATAGATGTGATCAGTGTGACACTTTCGAGTACATGGCAG	751
Qy	889	ACTCCGGGAGCCCCAGATCGCGGTTTCGACTGGAGCTGGTGTTCACGTGGCACAAG	948
Db	752	GTTCTGACATGACCTA---TGGCGGGTCAACTGGAAAGTCAACTTTCGCTGGTATCCTG	808
Qy	949	TTCTTGAGAGGAGAGG---ATACGGATGCAATCCCCGTGCGATGTGATCAGGTCTCAA	1005
Db	809	TTCCCCAAAGAGAAATGCAGAGAAAGGATGATCGGACTCTTCTCTGTGAGAAACACTA	868
Qy	1006	CAATGGCTGTGGGCTGTTGCTGTGAGTAGAATAATTTTGAATATCTGGGGTCTTATG	1065
Db	869	CAATGGCAGAGGCGCTTTTTCATAGACAGAGATTACTTTCAGGAATTTGAAACATATG	928
Qy	1066	ATACAGGAATCGAAGTTTGGGGAGAGAAAACCTCGAATTTTCCTTTAGGATCTGGCAGT	1125
Db	929	ATGCTGAATGGATATTTGGGGAGAGAAACCTAGAAATTTCTTTAGGATTTGGCAGT	988
Qy	1126	GTGGTGGGGTTTCGAAACACACCCATGTTCCCATGTGTGGCCATGTGTTTTCCCAAGCAAG	1185
Db	989	GTGAGAGAACTTTGGAGATGTACTTGTCTCACATGTTGGACATGTGTTTCGGAAGCTA	1048
Qy	1186	CTCCCTACTCCGCAACAAGGCTCTGGCCACAGTG-----TGTCGAC	1230
Db	1049	CACCTTACAGTTTTCAGGAGGCGCGGGCAGATTATCAATAAAAAATAACAGACGACTTG	1108
Qy	1231	CTGAAGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCGCCGCT	1290
Db	1109	CAGAAATATGATGGATGAATTCAGGAATTTCTCTATATAATTTCTCCAGGTGTTACAA	1168
Qy	1291	TGGAACCTTTTGGGATGTGCAGAGAGGAAGCAGCTCCGGGCAAACTCCAGTGTAAAG	1350
Db	1169	AGGTAGATTATGGAGATATATATCAAGACTTGGTCTAAGGSCAATACTCCAATGCGAC	1228
Qy	1351	ACTTCAAGTGGTCTTGGAGACTGTGATCCAGAACTGCATGTGCCCTGAGGACAGGCGTG	1410
Db	1229	CATTCTCTTGGTACCTAGAGAAATTTATCCTGATTCAGATTCTCTCGCTCATTTTCT	1288
Qy	1411	GCTTCTTCGGGATGCTCCAGAAACAAGGACTAACAGACTACTGCTTTTGACTATAACCTC	1470
Db	1289	CTTTGGAGAGATAC-----GAATGTGAAACAATACTAGTGTCTAGATAACATGGCTA	1342
Qy	1471	CCGATGAAAAACCAAGATTGTGGGACACAGGTCATTCTGTACTCTGTATGGGATGGGCC	1530
Db	1343	GAAGAAGAGATGAAGAAGTTGGA-----ATTTTAACTGTCTGTTATGGGAG	1390
Qy	1531	AGAACTAGTTTTTCAGTACACGCTCCAGAAAGAAAT	1567
Db	1391	GTAACTAGGTTTTTCTCTTACACTGCCAACCAAGAAAT	1427

RESULT 13
 US-08-967-508-10
 ; Sequence 10, Application US/08967508
 ; Patent No. 5910570
 ; GENERAL INFORMATION:
 ; APPLICANT: The Upjohn Company
 ; APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ak
 ; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-
 ; TITLE OF INVENTION: Polypeptide, N-Acetylgalact
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pharmacia & Upjohn Company, Intelle
 ; ADDRESSEE: Property Legal Services
 ; STREET: 301 Henrietta Street
 ; CITY: Kalamazoo
 ; STATE: Michigan
 ; COUNTRY: USA
 ; ZIP: 49001
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/967,508
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/602,830
 ; FILING DATE: 13 NO. 5910570ember 1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Barnley Jr., James D.
 ; REGISTRATION NUMBER: 33,673
 ; REFERENCE/DOCKET NUMBER: 4755.P CP
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 616-833-2210
 ; TELEFAX: 616-833-8897
 ; TELEX: 224401
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2294 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cdna
 US-08-967-508-10

Query Match	8.1%; Score 232.2; DB 2; Length 2294;
Best Local Similarity	54.8%; Pred. No. 6.1e-40;
Matches	612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;
Qy	472 ATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATTAATGAAGCCTGGTCAA 531
Db	384 ATCCAGATAACCTTCTTACAAACCAAGTGTGGTGAATGTTTTCACAAATGAGGCTTGGAGCA 443
Qy	532 CTCTCCTTCGACAGATTTACAGTGTCTCTTCAGACATCCCGGATATCTCTGTAGAGAG 591
Db	444 CACTTCTGCCAACTGTCCATAGCGTCAATTAATCGCTCACCAGGACATGCTTAGAAGAAA 503
Qy	592 TGATCCTTGTAGATGACTACAGTGATAGAGACACCTGAAGGAGCGCTTTGSCCAATGAGC 651
Db	504 TTGTTCTAGTAGATGATGCCAGTGAAGAGACCTTTTAAAGACCTCTTAGAGAGTTACG 563
Qy	652 TTTTCGGGACT---GCCCAAGGTGGCCTGATCCGCGCCAAACAAGAGAGAGGCGCTTGGTGC 708
Db	564 TGAATAAATTAAGAATACCGTTCACGTCAITCGAATGGAGCAGCGTTCTGGATTGATCA 623
Qy	709 GAGCCCGGCTGCTGGGGCGCTCTCGCGGAGGGGCGATGTTCTGACCTTCTCGGACTGTC 768
Db	624 GAGCTAGTTAAAGGTGCTGCTGTGTTCTAAAGGCCAAGTATCACCTTTTATAGCGCGC 683
Qy	769 ACTGTAGTCCCAAGAAAGGTGGCTGGAGCGCGTGTGTCAGAGGATCCATGAAGAGGAGT 828

Db 1281 CATTCTCTGTGCTAGAGAAATATTTATCTGATTCCTCAGATTCCTCGTCACTATTCT 1340
 Qy 1411 GCTTCTTCGGGATGCTCCAGNACAAAGGACTACAGACTACTGCTTTGACTATAACCTC 1470
 Db 1341 CTTTGGGAGAGATAC-----GAAATGTGGAACAAATCAGTGTCTAGATAACATGGCTA 1394
 Qy 1471 CCGATGAAACACGATTGTGGGACACCAGGTCAATCTGTACCTCTGTATGGGATGGGCC 1530
 Db 1395 GAAAGAGATGAAAGTTGGA-----ATTTTAACTGTCTATGGGAG 1442
 Qy 1531 AGAATCAGTTTTTCGAGTACACGTCGCCAGAAAGAAAT 1567
 Db 1443 GTAATCAGGTTTTCTCTTACACTGCCAACAAAGAAAT 1479

Search completed: December 7, 2003, 21:21:29
 Job time : 180.847 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 18:38:04 ; Search time 867.526 Seconds
(without alignments)
10918.700 Million cell updates/sec

Title: US-10-074-527-1
Perfect score: 2850
Sequence: 1 cggcgctggtaccactataa.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2848	99.9	2850	13	US-10-074-527-1
2	2802.4	98.3	2904	14	US-10-198-846-9749
3	1746	61.3	1746	13	US-10-074-527-3
4	1744.4	61.2	1746	12	US-10-292-896-57
5	1208.4	42.4	1259	9	US-09-925-299-219
6	1208.4	42.4	1259	11	US-09-925-299-219
7	1012.2	35.5	1517	14	US-10-106-698-598
8	1011.8	35.5	1517	11	US-09-895-298-44
9	584.6	20.5	818	14	US-10-198-846-7621
10	526.4	18.5	544	9	US-09-777-564-414
11	526.4	18.5	544	14	US-10-015-219-414
12	524	18.4	525	12	US-10-292-896-121
13	496.4	17.4	632	9	US-09-777-564-172
14	496.4	17.4	632	9	US-09-777-564-1423
15	496.4	17.4	632	14	US-10-015-219-172
16	496.4	17.4	632	14	US-10-015-219-1423

17	450.8	15.8	496	10	US-09-867-701-4790	Sequence 4790, Ap
c 18	443.4	15.6	473	10	US-09-867-701-4789	Sequence 4789, Ap
c 19	425.8	14.9	473	10	US-09-867-701-3683	Sequence 3683, Ap
c 20	409	14.4	421	10	US-09-967-768A-21	Sequence 21, Appl
c 21	409	14.4	421	10	US-09-873-367C-550	Sequence 550, App
c 22	406.4	14.3	408	10	US-09-867-701-10399	Sequence 10399, A
c 23	285.6	10.0	1920	9	US-09-789-417-1	Sequence 1, Appli
c 24	285.6	10.0	1920	9	US-09-792-451-1	Sequence 1, Appli
c 25	285.6	10.0	1920	10	US-10-292-896-59	Sequence 59, Appl
c 26	285.6	10.0	2575	11	US-09-946-374-346	Sequence 346, App
c 27	285.6	10.0	2575	12	US-10-015-387A-346	Sequence 346, App
c 28	285.6	10.0	2575	12	US-10-006-130A-346	Sequence 346, App
c 29	285.6	10.0	2575	12	US-10-199-672-363	Sequence 363, App
c 30	285.6	10.0	2575	12	US-10-006-172A-346	Sequence 346, App
c 31	285.6	10.0	2575	12	US-10-187-749-363	Sequence 363, App
c 32	285.6	10.0	2575	12	US-10-194-457-363	Sequence 363, App
c 33	285.6	10.0	2575	12	US-10-184-642-363	Sequence 363, App
c 34	285.6	10.0	2575	12	US-10-196-747-363	Sequence 363, App
c 35	285.6	10.0	2575	12	US-10-015-392A-346	Sequence 346, App
c 36	285.6	10.0	2575	12	US-10-017-253A-346	Sequence 346, App
c 37	285.6	10.0	2575	12	US-10-173-689-363	Sequence 363, App
c 38	285.6	10.0	2575	12	US-10-173-690-363	Sequence 363, App
c 39	285.6	10.0	2575	12	US-10-173-691-363	Sequence 363, App
c 40	285.6	10.0	2575	12	US-10-173-692-363	Sequence 363, App
c 41	285.6	10.0	2575	12	US-10-173-694-363	Sequence 363, App
c 42	285.6	10.0	2575	12	US-10-173-698-363	Sequence 363, App
c 43	285.6	10.0	2575	12	US-10-173-699-363	Sequence 363, App
c 44	285.6	10.0	2575	12	US-10-173-707-363	Sequence 363, App
c 45	285.6	10.0	2575	12	US-10-174-569-363	Sequence 363, App

ALIGNMENTS

RESULT 1

US-10-074-527-1
; Sequence 1, Application US/10074527
; Publication No. US20020142426A1
; GENERAL INFORMATION:
; APPLICANT: Olandt, Peter J.
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Galvin, Katherine A.
; APPLICANT: Millennium Pharmaceuticals Inc.
; TITLE OF INVENTION: 33945, A Human Glycosyltransferase and
; FILE REFERENCE: MPI2001-018P1RCPI(M)
; CURRENT APPLICATION NUMBER: US/10/074,527
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/269202
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2850
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (81)...(1826)
US-10-074-527-1

Query Match 99.9%; Score 2848; DB 13; Length 2850;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2850; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCGGCTCGGTACCACTATAACGCCGCCAGTGTGCGAATTCGCCCTTCGCAGATCGC 60
Db 1 CCGGCTCGGTACCACTATAACGCCGCCAGTGTGCGAATTCGCCCTTCGCAGATCGC 60
Qy 61 TGGCTCGAGTTCGGCGGCGCATGTGGGGGCGCACCGCGCGCGCTGCCCGGGAAC 120
Db 61 TGGCTCGAGTTCGGCGGCGCATGTGGGGGCGCACCGCGCGCGCTGCCCGGGAAC 120


```
Db 2281 GGGTTAAAGTTTCCCAAGTATAGAGAGACTGTCTACTAGGACATTGTATTGATTATTC 2340
Qy 2341 GGTCAITGAGATCTCTAGATGTAATTTTAAAGAAAGTCTTTTGGTTATGTGTCTAC 2400
Db 2341 GGTCAITGAGATCTCTAGATGTAATTTTAAAGAAAGTCTTTTGGTTATGTGTCTAC 2400
Qy 2401 CACAGTTAACTCCATTAATGTTTCATGTGAGCCAAAGAGAGACTAACAAAGCTGAAATCT 2460
Db 2401 CACAGTTAACTCCATTAATGTTTCATGTGAGCCAAAGAGAGACTAACAAAGCTGAAATCT 2460
Qy 2461 CAGAGAACTAATTTGCTTTTACTAGCTGAGTCAACTTCAGAGCGAACTTCTTAACTGCGG 2520
Db 2461 CAGAGAACTAATTTGCTTTTACTAGCTGAGTCAACTTCAGAGCGAACTTCTTAACTGCGG 2520
Qy 2521 CACTGTAGTGGCTGTTCTTACCACTATGACTTTTAAACATGTTTATATCAITTTTAAAT 2580
Db 2521 CACTGTAGTGGCTGTTCTTACCACTATGACTTTTAAACATGTTTATATCAITTTTAAAT 2580
Qy 2581 TTTTATGATACGGTAGTGTGAGGAGAAATGTAATGTTCTATATGAAATTCCTTTTCAA 2640
Db 2581 TTTTATGATACGGTAGTGTGAGGAGAAATGTAATGTTCTATATGAAATTCCTTTTCAA 2640
Qy 2641 GTTTGTTTCAATTAATACAGTTTATTAATTTAAATCAGCGTTAGAGTTTGTCTGCTGCAAC 2700
Db 2641 GTTTGTTTCAATTAATACAGTTTATTAATTTAAATCAGCGTTAGAGTTTGTCTGCTGCAAC 2700
Qy 2701 TGCTGTGAAATTTCTCTGAGTAAATCTGATTTTGTGAATGATCCAGACCAACCTCTGAGA 2760
Db 2701 TGCTGTGAAATTTCTCTGAGTAAATCTGATTTTGTGAATGATCCAGACCAACCTCTGAGA 2760
Qy 2761 TTTTGTCAACCTGATTAAGTCAATATGAATGATTTAAAGATGTGAGAACAAAAA 2820
Db 2761 TTTTGTCAACCTGATTAAGTCAATATGAATGATTTAAAGATGTGAGAACAAAAA 2820
Qy 2821 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2850
Db 2821 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2850
```

```
RESULT 2
US-10-198-846-9749/c
; Sequence 9749, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steimann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9749
; LENGTH: 2904
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 3, 4, 5, 2903, 2904
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-9749
```

```
Query Match 98.3%; Score 2802.4; DB 14; Length 2904;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2811; Conservative 1; Indels 1; Gaps 1;
Qy 12 CCACTATAACGGCGCGCAGTGTCTGGAATTCGCCCTTGGCAGATGCTGGCTGCAGTT 971
```

```
Db 2894 CCACTATAACGGCGCGC -GTGTCTGGAATTCGCCCTTGGCAGATGCTGGCTGCAGTT 2836
Qy 72 GCGGGCGCATGTGGGGGCGCACGGCGCGCGCTGCCCGGGGAACCTCGCGCGCGC 131
Db 2835 GCGGGCGCATGTGGGGGCGCACGGCGCGCGCTGCCCGGGGAACCTCGCGCGCGC 2776
Qy 132 CGGAGGCGCTGTGTGTCTCTGGGGCTACTGTGGCTTGGCCGGGCTGGGCTCGTGTCTG 191
Db 2775 CGGAGGCGCTGTGTGTCTCTGGGGCTACTGTGGCTTGGCCGGGCTGGGCTCGTGTCTG 2716
Qy 192 CGGCGCAGGCTGGGGCGGGGCGGGGCTGCCAGCGCGGACCCCGCGCACCCCGCGC 251
Db 2715 CGGCGCAGGCTGGGGCGGGGCGGGGCTGCCAGCGCGGACCCCGCGCACCCCGCGC 2656
Qy 252 CCGGCGCGCGCGCAGCGGCTCATGCGCGCGCGCGCTGCCAGCGCGGACCCCGCGC 311
Db 2655 CCGGCGCGCGCGCAGCGGCTCATGCGCGCGCGCGCTGCCAGCGCGGACCCCGCGC 2596
Qy 312 CGGCGCGAGCGGCTGGGGCTGCGAGCTGCGAGGCGAGGAGCTGGGGCTGCGAGGAGAGC 371
Db 2595 CGGCGCGAGCGGCTGGGGCTGCGAGCTGCGAGGCGAGGAGCTGGGGCTGCGAGGAGAGC 2536
Qy 372 GTGGGCTGCACAGATTAACATCTACTCAGCGACCGCATCTCACTGCAACCGCGCGCTG 431
Db 2535 GTGGGCTGCACAGATTAACATCTACTCAGCGACCGCATCTCACTGCAACCGCGCGCTG 2476
Qy 432 CCGGAGCGCTGGAAACCGCTGTGCAAGAGAGAAATATGATTAATGATTAATTTGCCAGG 491
Db 2475 CCGGAGCGCTGGAAACCGCTGTGCAAGAGAGAAATATGATTAATGATTAATTTGCCAGG 2416
Qy 492 ACATCTGTTATCATAGCATTTTATAATGAAGCCTGGTCAACTCTCTTCGGAAGTTTAC 551
Db 2415 ACATCTGTTATCATAGCATTTTATAATGAAGCCTGGTCAACTCTCTTCGGAAGTTTAC 2356
Qy 552 AGTGTCTTGAAGACATCCCGGATATCTGCTAGAGAGAGATGATTCCTTGTAGATGACTAC 611
Db 2355 AGTGTCTTGAAGACATCCCGGATATCTGCTAGAGAGAGATGATTCCTTGTAGATGACTAC 2296
Qy 612 AGTGTATAGAGACCATGGAAGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGGTG 671
Db 2295 AGTGTATAGAGACCATGGAAGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCAAGGTG 2236
Qy 672 CGCTGATCCGCGCCAAAGAGAGAGGCGCTGCTGCGAGCGCGGCTGCTGGGGGCGCTCT 731
Db 2235 CGCTGATCCGCGCCAAAGAGAGAGGCGCTGCTGCGAGCGCGGCTGCTGGGGGCGCTCT 2176
Qy 732 GCGCGAGGGGCGATGTTCTGACCTTCTGGACTGTCACTGTAGTGCCCAAGAGGCTG 791
Db 2175 GCGCGAGGGGCGATGTTCTGACCTTCTGGACTGTCACTGTAGTGCCCAAGAGGCTG 2116
Qy 792 CTGGAGCGCTGCTGAGAGGATCCATGAAGAGGAGTCCGAGTGGTGGTCCCGGCTGATT 851
Db 2115 CTGGAGCGCTGCTGAGAGGATCCATGAAGAGGAGTCCGAGTGGTGGTCCCGGCTGATT 2056
Qy 852 GATGTGATCGACTGGAACACCTTCGAATACCTGGGAACTCCCGGAGCGCCAGATCGGC 911
Db 2055 GATGTGATCGACTGGAACACCTTCGAATACCTGGGAACTCCCGGAGCGCCAGATCGGC 1996
Qy 912 GGTTCGACTGAGAGGCTGTGTTTCAGTGGCACTGTTCTGAGAGGAGAGGATACG 971
Db 1995 GGTTCGACTGAGAGGCTGTGTTTCAGTGGCACTGTTCTGAGAGGAGAGGATACG 1936
Qy 972 ATGCAATCCCGCTGCTGATCTCAGTCTCCAAACAAATGGCTGGCTGGTGTGCTGTG 1031
Db 1935 ATGCAATCCCGCTGCTGATCTCAGTCTCCAAACAAATGGCTGGCTGGTGTGCTGTG 1876
Qy 1032 AGTAAGAAATATTTTGAATATCTGGGCTCTTATGATACAGGAATGAAAGTTTGGGAGGA 1091
Db 1875 AGTAAGAAATATTTGAATATCTGGGCTCTTATGATACAGGAATGAAAGTTTGGGAGGA 1816
Qy 1092 GAAAACTCGAAATTTTCTTTTAGGATCTGGCAGTGTGGTGGGCTTCTGGAAACACACCCA 1151
```

1815 GAAACCTCGAATTTTCTCTTTAGGATCTGGCAGTGTGGTGGGGTTCTGGAAACACACCCA 1756
1152 TGTTCCTCATGTTGGCCATGTTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTG 1211
1755 TGTTCCTCATGTTGGCCATGTTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTG 1696
1212 GCCAACAGTGTCTGTGAGCTGAAGTATGATGATGAAATTTAAAGAGCTCTACTACCAT 1271
1695 GCCAACAGTGTCTGTGAGCTGAAGTATGATGATGAAATTTAAAGAGCTCTACTACCAT 1636
1272 CGCAACCCCGTGGCCGCTTGGACCTTTTGGGATGTGACAGAGGAAAGCACTCCGG 1331
1635 CGCAACCCCGTGGCCGCTTGGACCTTTTGGGATGTGACAGAGGAAAGCACTCCGG 1576
1332 GACAGCTCCAGTGAAGACTTCAAGTGTCTTGGAGTGTGATCCAGAACTGCAT 1391
1575 GACAAGCTCCAGTGAAGACTTCAAGTGTCTTGGAGTGTGATCCAGAACTGCAT 1516
1392 GTGCTCGAGACAGGCTGCTTCTCGGATGCTCCAGAACAAAGGACTTAAACAGACTAC 1451
1515 GTGCTCGAGACAGGCTGCTTCTCGGATGCTCCAGAACAAAGGACTTAAACAGACTAC 1456
1452 TGCTTTGACTATACCTCCCGATGAAACAGATTGTGGGACACCAAGGTCATTCTGTAC 1511
1455 TGCTTTGACTATACCTCCCGATGAAACAGATTGTGGGACACCAAGGTCATTCTGTAC 1396
1512 CTGCTCATGGATGGGCCAGAAATCAGTTTTCAGTACAGCTCCAGAAAGAAATACGC 1571
1395 CTGCTCATGGATGGGCCAGAAATCAGTTTTCAGTACAGCTCCAGAAAGAAATACGC 1336
1572 TATAACACCCACAGCTGAGGCTGCAATGCTGGAAGCAGGAATGGATACCTTATC 1631
1335 TATAACACCCACAGCTGAGGCTGCAATGCTGGAAGCAGGAATGGATACCTTATC 1276
1632 ATGCACTCTGCGAAGAACTGCCCCAGAGAAATCAGAAATTCATCTTCAGGAGGATGGA 1691
1275 ATGCACTCTGCGAAGAACTGCCCCAGAGAAATCAGAAATTCATCTTCAGGAGGATGGA 1216
1692 TCTTTATTTCAGAACTGCAAGAAATGCTCCAGCTGGGAGGAGGATCGAGTAC 1751
1215 TCTTTATTTCAGAACTGCAAGAAATGCTCCAGCTGGGAGGAGGATCGAGTAC 1156
1752 AGTTTCGTTCCACTCTTACGAGACTGCACCAACTCGGATCATCAAGAAATGTTCTTCAAA 1811
1155 AGTTTCGTTCCACTCTTACGAGACTGCACCAACTCGGATCATCAAGAAATGTTCTTCAAA 1096
1812 GAGCGCATGTTATGAAGCTCGTGTATCAAGAGGCCCATCGAAGGAGACTGTGAGCCAG 1871
1095 GAGCGCATGTTATGAAGCTCGTGTATCAAGAGGCCCATCGAAGGAGACTGTGAGCCAG 1036
1872 GACTCTGCCCAACAAAGACTTAGCTAAGCAGTACCAAGACCCCAACCAAACTAGGCTGC 1931
1035 GACTCTGCCCAACAAAGACTTAGCTAAGCAGTACCAAGACCCCAACCAAACTAGGCTGC 976
1932 ATTCTTTGAAGAGCAATCAATTTTGCATTTGGAAGTGTGCTGGATTAGTAAAAA 1991
975 ATTCTTTGAAGAGCAATCAATTTTGCATTTGGAAGTGTGCTGGATTAGTAAAAA 916
1992 TGTGAATAAGCTTTGTACTTTATTTTGAGAACTTTTAAATGTTCCAAATAACCTATTTT 2051
915 TGTGAATAAGCTTTGTACTTTATTTTGAGAACTTTTAAATGTTCCAAATAACCTATTTT 856
2052 CAAAGGGTAATCGTAAGTGTATACCTTGTATTTAGAAATTAACCTTATAATTT 2111
855 CAAAGGGTAATCGTAAGTGTATACCTTGTATTTAGAAATTAACCTTATAATTT 796
2112 TTTCTATCAARAWAWATTTTACAGTCGTGCTTTTACTCTCATTTAGCAAAAAAGATAA 2171
795 TTTCTATCAAGATGTATATTTTACAGTCGTGCTTTTACTCTCATTTAGCAAAAAAGATAA 736
2172 AGATTTTATTTTGTATTTTACAAGAAATTTCCAGGTACGAAGATATCTGATGGGTGAAA 2231
735 AGATTTTATTTTGTATTTTACAAGAAATTTCCAGGTACGAAGATATCTGATGGGTGAAA 676

2232 TCAGGTTCAAGCAACGTACTTTTGCATTAACATGATAATACCTCAGCTGCGGGTTAAAGTT 2291
675 TCAGGTTCAAGCAACGTACTTTTGCATTAACATGATAATACCTCAGCTGCGGGTTAAAGTT 616
2292 TTCCCAAGTATAGAGAGACTGTCACTAGGAACTATTGTAATTTATTCAGGTCAATTGAGA 2351
615 TTCCCAAGTATAGAGAGACTGTCACTAGGAACTATTGTAATTTATTCAGGTCAATTGAGA 556
2352 TCTTCTAGATGTATTTTAAAGAAATGCTTTTGGTATGTTGTTGCTACCAAGTTAAACA 2411
555 TCTTCTAGATGTATTTTAAAGAAATGCTTTTGGTATGTTGTTGCTACCAAGTTAAACA 496
2412 CTCCTAATATGTTCTATGCTCAGCCAAAGAGGACTTAAACCAAGCTGAAATCTCAGAGAACAAT 2471
495 CTCCTAATATGTTCTATGCTCAGCCAAAGAGGACTTAAACCAAGCTGAAATCTCAGAGAACAAT 436
2472 TTGCTTTTACTAAGCTCAGTCAACTTGAGAGCGAACTTCTAAACAAATGCCGACCTGTAGTGT 2531
435 TTGCTTTTACTAAGCTCAGTCAACTTGAGAGCGAACTTCTAAACAAATGCCGACCTGTAGTGT 376
2532 GGCTGTTTCTACCACTATGACTTTAAACATGTTTATATCATTTTAAATTTTATGATAC 2591
375 GGCTGTTTCTACCACTATGACTTTAAACATGTTTATATCATTTTAAATTTTATGATAC 316
2592 GGCTGTTTCTACCACTATGACTTTAAATGTTTATATGAAATTCCTTTTCAAGTTTGTTCATT 2651
315 GGCTGTTTCTACCACTATGACTTTAAATGTTTATATGAAATTCCTTTTCAAGTTTGTTCATT 256
2652 AATAACAGTTTATTAATTTAAATCAGCGTTAGAGTTTGTGCTGCTGCAACTGTGCTGAAAA 2711
255 AATAACAGTTTATTAATTTAAATCAGCGTTAGAGTTTGTGCTGCTGCAACTGTGCTGAAAA 196
2712 TTTCTCTGAGTAATTTCTGATTTTGTGATGATCCAGACCAACCTCGAGATTTTGTCAACC 2771
195 TTTCTCTGAGTAATTTCTGATTTTGTGATGATCCAGACCAACCTCGAGATTTTGTCAACC 136
2772 TGATTAAGTCAATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2829
135 TGATTAAGTCAATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 78

RESULT 3

US-10-074-527-3
; Sequence 3, Application US/10074527
; Publication No. US20020142426A1
; GENERAL INFORMATION:
; APPLICANT: Olandt, Peter J.
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Galvin, Katherine A.
; APPLICANT: Millennium Pharmaceuticals Inc.
; TITLE OF INVENTION: 33945, A Human Glycosyltransferase and
; TITLE OF INVENTION: Uses therefor
; FILE REFERENCE: MPI2001-018P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/074, 527
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/269202
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1746
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1746)
US-10-074-527-3

Query Match 61.3%; Score 1746; DB 13; Length 1746;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1746; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 ATGTGGGGCGCAGCGCGCGCGCTGCCCGGGGAACTGGCGCGCGCGCGCGGAGCG 140
Db 1 ATGTGGGGCGCAGCGCGCGCGCGCTGCCCGGGGAACTGGCGCGCGCGCGCGGAGCG 60
QY 141 CTGTTGTTGCTCTCTGGCGCTACTGCGCTTGGCGGGGCTGGGCTCGGTGCTGCGGGCGCAG 200
Db 61 CTGTTGTTGCTCTCTGGCGCTACTGCGCTTGGCGGGGCTGGGCTCGGTGCTGCGGGCGCAG 120
QY 201 CGTGGGGCGGGCGCGGGCTGCCAGCGCGGGACCCCGCGGCACTCCCGCGCGCGCGCGCGG 260
Db 121 CGTGGGGCGGGCGCGGGCTGCCAGCGCGGGACCCCGCGGCACTCCCGCGCGCGCGCGG 180
QY 261 CGCGAGCGGTTCATGCGCGCGCGCGCGGTGCGCGGAACTGGCGCGCGCGCGCGGCGAG 320
Db 181 CGCGAGCGGTTCATGCGCGCGCGCGCGGTGCGCGGAACTGGCGCGCGCGCGCGGCGAG 240
QY 321 GCGGTGCGGTGTCAGCTGTCAGGGCGAGGAGCTGCGGCTGCGAGGAGAGAGCGTGCAGCTG 380
Db 241 GCGGTGCGGTGTCAGCTGTCAGGGCGAGGAGCTGCGGCTGCGAGGAGAGAGCGTGCAGCTG 300
QY 381 CACAGATTAACTTACCTCAGCGACCGCATCTCACTGACACCGCGCGCTGCCCGAGCGC 440
Db 301 CACAGATTAACTTACCTCAGCGACCGCATCTCACTGACACCGCGCGCTGCCCGAGCGC 360
QY 441 TGGNACCGCTGTGCAAGAGAGAAATATATATATATATATATATATATATATATATATATAT 500
Db 361 TGGNACCGCTGTGCAAGAGAGAAATATATATATATATATATATATATATATATATATATAT 420
QY 501 ATCATAGCATTTTATATAGAGCTGTGCACTCTCTCGGACAGTTTACAGTGTCTCTT 560
Db 421 ATCATAGCATTTTATATAGAGCTGTGCACTCTCTCGGACAGTTTACAGTGTCTCTT 480
QY 561 GAGACATCCCCGGATATCTCTAGAGAGAGTATCTCTGTAGATGACTACAGTGATAGA 620
Db 481 GAGACATCCCCGGATATCTCTAGAGAGAGTATCTCTGTAGATGACTACAGTGATAGA 540
QY 621 GAGCACTGAGGAGCGCTTGGCCAACTGAGCTTTCGGGACTGCCAAGGTGCGCTGATC 680
Db 541 GAGCACTGAGGAGCGCTTGGCCAACTGAGCTTTCGGGACTGCCAAGGTGCGCTGATC 600
QY 681 CGCGCCAAAGAGAGAGGCGCTGTGCGAGCGCGCGCTGCGGGGCGCTGCGGGCGAGG 740
Db 601 CGCGCCAAAGAGAGAGGCGCTGTGCGAGCGCGCGCTGCGGGGCGCTGCGGGCGAGG 660
QY 741 GCGCATGTTCTGACCTTCTCTGAGCTGTCACTGTGAGTGCCACGAAGGTGCGCTGAGC 800
Db 661 GCGCATGTTCTGACCTTCTCTGAGCTGTCACTGTGAGTGCCACGAAGGTGCGCTGAGC 720
QY 801 CTGCTGAGAGGATCCATGAAGAGAGTGGCAGTGTGTCGCGGTGATGTATGTATC 860
Db 721 CTGCTGAGAGGATCCATGAAGAGAGTGGCAGTGTGTCGCGGTGATGTATGTATC 780
QY 861 GACTTGAACACCTTCGAATACCTGGGAACTCCGGGAGCGCCAGATCGCGGTTCGAC 920
Db 781 GACTTGAACACCTTCGAATACCTGGGAACTCCGGGAGCGCCAGATCGCGGTTCGAC 840
QY 921 TGGAGGTGCTGTTCAGTGGGCAACAGTTTCTGAGAGGAGAGGATACGGATGCAATCC 980
Db 841 TGGAGGTGCTGTTCAGTGGGCAACAGTTTCTGAGAGGAGAGGATACGGATGCAATCC 900
QY 981 CCGGTGATGTCATCAGGTCTCAACAAATGGCTGGTGGGCTGTTGCTGTGAGTAAAGAA 1040
Db 901 CCGGTGATGTCATCAGGTCTCAACAAATGGCTGGTGGGCTGTTGCTGTGAGTAAAGAA 960
QY 1041 TATTTTGAATATCGGGTCTTATGATACAGGAATGGAAGTTTGGGAGAGAGAAACCTC 1100
Db 961 TATTTTGAATATCGGGTCTTATGATACAGGAATGGAAGTTTGGGAGAGAGAAACCTC 1020
QY 1101 GAATTTTCTTTAGGATCTGGCAGTGTGGTGGTCTCTGGAACACACCCATGTTCCTCAT 1160
Db 1021 GAATTTTCTTTAGGATCTGGCAGTGTGGTGGTCTCTGGAACACACCCATGTTCCTCAT 1080
QY 1161 GTTGGGCATGTTTTCCCAAGAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGT 1220

RESULT 4

US-10-292-896-57

; Sequence-57 Application US/10292896

; Publication No. US20030186850A1

; GENERAL INFORMATION:

; APPLICANT: HASSAN, Helle

; APPLICANT: REIS, Celso A.

; APPLICANT: BENNETT, Eric P.

; APPLICANT: CLAUSEN, Henrik

; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GAI

; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS

; FILE REFERENCE: 4305/1H154-US3

; CURRENT APPLICATION NUMBER: US/10/292,896

; CURRENT FILING DATE: 2003-03-17

; PRIOR APPLICATION NUMBER: US 60/425,204

; PRIOR FILING DATE: 2002-11-08

; PRIOR APPLICATION NUMBER: PCT/DK01/00328

; PRIOR FILING DATE: 2001-05-10

; PRIOR APPLICATION NUMBER: US 60/203,331

; PRIOR FILING DATE: 2000-05-11

; NUMBER OF SEQ ID NOS: 130

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 57

; LENGTH: 1746


```
; LENGTH: 1259
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-299-219

Query Match      42.4%; Score 1208.4; DB 9; Length 1259;
Best Local Similarity 98.4%; Pred. No. 6.6e-270;
Matches 1226; Conservative 4; Mismatches 15; Indels 1; Gaps 1;

Qy 292 CGCGAAACGCGCTGGGCGCGCGGGCGAGGCGGCTGCGCTGCGAGCTGCGAGGCGGAGGAGC 351
Db 7 CGGTGCCGGAACCGCTGGGCGGAGGAGGCGGCTGCGGCTGCGAGCTGCGAGGCGGAGGAGC 66

Qy 352 TCGGCTTCAGAGAGAGAGAGCTGCGGCTGCGACCGAGATTAACTTACCTCAGGACCGCA 411
Db 67 TCGGCTTCAGAGAGAGAGAGCTGCGGCTGCGACCGAGATTAACTTACCTCAGGACCGCA 126

Qy 412 TCTCACTGCACCGCGCTGCGGCGCGCTGCGGAGCGCTGCGAACCCTGCTGCAAGAGAGAAATATG 471
Db 127 TCTCACTGCACCGCGCTGCGGCGCGCTGCGAACCCTGCTGCAAGAGAGAAATATG 186

Qy 472 ATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCCCTGGTCAA 531
Db 187 ATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCCCTGGTCAA 531

Qy 532 CTCTCTTCGGACAGTTTACAGTGTCTTGTAGACATCCCGGATATCTCTGCTAGAGAGAG 591
Db 247 CTCTCTTCGGACAGTTTACAGTGTCTTGTAGACATCCCGGATATCTCTGCTAGAGAGAG 306

Qy 592 TGATCCTTGTAGATGACTACAGTGATAGAGAGACCTTGAAGAGCGCTTGGCCCAATGAGC 651
Db 307 TGATCCTTGTAGATGACTACAGTGATAGAGAGACCTTGAAGAGCGCTTGGCCCAATGAGC 366

Qy 652 TTTCCGAGCTCCCAAGGTGCGCTGATGCCGCGCAACAGAGAGAGGCGCTGCGGAG 711
Db 367 TTTCCGAGCTCCCAAGGTGCGCTGATGCCGCGCAACAGAGAGAGGCGCTGCGGAG 426

Qy 712 CCCGGCTGCTGGGCGCTGCTGCGGAGGCGGATGTTCTGACCTTCTGCACTGTCACT 771
Db 427 CCCGGCTGCTGGGCGCTGCTGCGGAGGCGGATGTTCTGACCTTCTGCACTGTCACT 486

Qy 772 GTGAGTGCCAGAGGCTGCTGAGCGCTGCTGCGAGGATCCATGAAGAGAGTGG 831
Db 487 GTGAGTGCCAGAGGCTGCTGAGCGCTGCTGCGAGGATCCATGAAGAGAGTGG 545

Qy 832 CAGTGTGTGCGCGGTGATGATGATCGACCGCAACCTTCCAACTCTGCGGAACT 891
Db 546 CAGTGTGTGCGCGGTGATGATGATCGACCGCAACCTTCCAACTCTGCGGAACT 605

Qy 892 CCGGGAGCCCGAGATCGGCGGTTTCGACTGGAGGCTGGTTCACGTGGCACAAGTTC 951
Db 606 CCGGGAGCCCGAGATCGGCGGTTTCGACTGGAGGCTGGTTCACGTGGCACAAGTTC 665

Qy 952 CTGAGAGGAGAGATACGGATGCAATCCCGCTGATGCTGATGCTGATGCTGCAATATG 1011
Db 666 CTGAGAGGAGAGATACGGATGCAATCCCGCTGATGCTGATGCTGATGCTGCAATATG 725

Qy 1012 CTGCTGGGCTGTTTGTGAGTGAAGAAATATTTTGAATATCTGCGGTCTTATCATACAG 1071
Db 726 CTGCTGGGCTGTTTGTGAGTGAAGAAATATTTTGAATATCTGCGGTCTTATCATACAG 785

Qy 1072 GAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTG 1131
Db 786 GAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTG 845

Qy 1132 GGGTCTTGGAAACACACCCATGTTCCATGTTGGCCATGTTTCCCAAGCAAGCTCCCT 1191
Db 846 GGGTCTTGGAAACACACCCATGTTCCATGTTGGCCATGTTTCCCAAGCAAGCTCCCT 905

Qy 1192 ACTCCCGCAACAAGGCTCTGSCCAACAGTGTTCGTGAGCTGGAAGTATGATGATGAAT 1251
Db 906 ACTCCCGCAACAAGGCTCTGSCCAACAGTGTTCGTGAGCTGGAAGTATGATGATGAAT 965
```

```
Qy 1252 TTAAAGAGCTTACTACTACATCGCAACCCCGTGCCTTGGAAACCTTTTGGGATGTGA 1311
Db 966 TTAAAGAGCTTACTACTACATCGCAACCCCGTGCCTTGGAAACCTTTTGGGATGTGA 1025

Qy 1312 CAGAGAGAGAGAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGGTCTTGGAGA 1371
Db 1026 CAGAGAGAGAGAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGGTCTTGGAGA 1085

Qy 1372 CTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCTGGCTTCTTCGGGATGCTCCAGA 1431
Db 1086 CTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCTGGCTTCTTCGGGATGCTCCAGA 1145

Qy 1432 ACAAAGAGACTAAAGACTACTGCTTTGACTATTAACCTCCCGATGAAAAACAGATTGTGG 1491
Db 1146 ACAAAGAGACTAAAGACTACTGCTTTGACTATTAACCTCCCGATGAAAAACAGATTGTGG 1205

Qy 1492 GACACAGGCTCATCTGTACTCTGTCATGGATGGGCGCAGAAATCA 1537
Db 1206 GACACAGGCTCATCTGTACTCTGTCATGGGATGGGCGCAGAAATGA 1251

RESULT 6
US-09-925-299-219
; Sequence 219, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 219
; LENGTH: 1259
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-299-219

Query Match      42.4%; Score 1208.4; DB 11; Length 1259;
Best Local Similarity 98.4%; Pred. No. 6.6e-270;
Matches 1226; Conservative 4; Mismatches 15; Indels 1; Gaps 1;

Qy 292 CGCGAAACGCGCTGGGCGCGCGGGCGAGGCGGCTGCGCTGCGAGCTGCGAGGCGGAGGAGC 351
Db 7 CGGTGCCGGAACCGCTGGGCGGAGGAGGCGGCTGCGGCTGCGAGCTGCGAGGCGGAGGAGC 66

Qy 352 TCGGCTTCAGAGAGAGAGAGCTGCGGCTGCGACCGAGATTAACTTACCTCAGGACCGCA 411
Db 67 TCGGCTTCAGAGAGAGAGAGCTGCGGCTGCGACCGAGATTAACTTACCTCAGGACCGCA 126

Qy 412 TCTCACTGCACCGCGCTGCGGCGCGCTGCGGAGCGCTGCGAACCCTGCTGCAAGAGAGAAATATG 471
Db 127 TCTCACTGCACCGCGCTGCGGCGCGCTGCGAACCCTGCTGCAAGAGAGAAATATG 186

Qy 472 ATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCCCTGGTCAA 531
Db 187 ATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCCCTGGTCAA 246

Qy 532 CTCTCTTCGGACAGTTTACAGTGTCTTGTAGACATCCCGGATATCTCTGCTAGAGAGAG 591
Db 247 CTCTCTTCGGACAGTTTACAGTGTCTTGTAGACATCCCGGATATCTCTGCTAGAGAGAG 306

Qy 592 TGATCCTTGTAGATGACTACAGTGATAGAGAGACCTTGAAGAGCGCTTGGCCCAATGAGC 651
Db 307 TGATCCTTGTAGATGACTACAGTGATAGAGAGACCTTGAAGAGCGCTTGGCCCAATGAGC 366

Qy 652 TTTCCGAGCTCCCAAGGTGCGCTGATGCCGCGCAACAGAGAGAGGCGCTGCGGAG 711
Db 367 TTTCCGAGCTCCCAAGGTGCGCTGATGCCGCGCAACAGAGAGAGGCGCTGCGGAG 426

Qy 712 CCCGGCTGCTGGGCGCTGCTGCGGAGGCGGATGTTCTGACCTTCTGCACTGTCACT 771
Db 427 CCCGGCTGCTGGGCGCTGCTGCGGAGGCGGATGTTCTGACCTTCTGCACTGTCACT 486

Qy 772 GTGAGTGCCAGAGGCTGCTGAGCGCTGCTGCGAGGATCCATGAAGAGAGTGG 831
Db 487 GTGAGTGCCAGAGGCTGCTGAGCGCTGCTGCGAGGATCCATGAAGAGAGTGG 545

Qy 832 CAGTGTGTGCGCGGTGATGATGATCGACCGCAACCTTCCAACTCTGCGGAACT 891
Db 546 CAGTGTGTGCGCGGTGATGATGATCGACCGCAACCTTCCAACTCTGCGGAACT 605

Qy 892 CCGGGAGCCCGAGATCGGCGGTTTCGACTGGAGGCTGGTTCACGTGGCACAAGTTC 951
Db 606 CCGGGAGCCCGAGATCGGCGGTTTCGACTGGAGGCTGGTTCACGTGGCACAAGTTC 665

Qy 952 CTGAGAGGAGAGATACGGATGCAATCCCGCTGATGCTGATGCTGATGCTGCAATATG 1011
Db 666 CTGAGAGGAGAGATACGGATGCAATCCCGCTGATGCTGATGCTGATGCTGCAATATG 725

Qy 1012 CTGCTGGGCTGTTTGTGAGTGAAGAAATATTTTGAATATCTGCGGTCTTATCATACAG 1071
Db 726 CTGCTGGGCTGTTTGTGAGTGAAGAAATATTTTGAATATCTGCGGTCTTATCATACAG 785

Qy 1072 GAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTG 1131
Db 786 GAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTG 845

Qy 1132 GGGTCTTGGAAACACACCCATGTTCCATGTTGGCCATGTTTCCCAAGCAAGCTCCCT 1191
Db 846 GGGTCTTGGAAACACACCCATGTTCCATGTTGGCCATGTTTCCCAAGCAAGCTCCCT 905

Qy 1192 ACTCCCGCAACAAGGCTCTGSCCAACAGTGTTCGTGAGCTGGAAGTATGATGATGAAT 1251
Db 906 ACTCCCGCAACAAGGCTCTGSCCAACAGTGTTCGTGAGCTGGAAGTATGATGATGAAT 965
```


Db 367 TTTCCGAGCTCCCAAGGTGGCCCTGATCCGCGCAACAAGAGAGAGGGCTGTGGAG 426
Qy 712 CCGGCTGCTGGGGCGTCTGGGCGGGGCGATGTTCTGACCTTCTGAGCTGTCACT 771
Db 427 CCGGCTGCTGGGGCGTCTGGGCGGGGCGATGTTCTGACCTTCTGAGCTGTCACT 486
Qy 772 GTGAGTCCCAAGAGGTGGCTGGAGCCGCTGCTGCAGAGGATCCATGAAGAGAGTCGG 831
Db 487 GTGAGTCCCAAGAGGT-GCTGGAGCGCTGCTGCAGAGGATCCATGAAGAGAGTCGG 545
Qy 832 CAGTGGTGTCCCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 891
Db 546 CAGTGGTGTCCCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 605
Qy 892 CCGGGAGCCCCAGATCGGCGGTTCGACTCGAGGCTGGTGTTCACGTGGGCACACAGTTC 951
Db 606 CCGGGAGCCCCAGATCGGCGGTTCGACTCGAGGCTGGTGTTCACGTGGGCACACAGTTC 665
Qy 952 CTGAGAGGAGAGGATACGGATGCAATCCCCCGCTCGATGTCATCAGGTCTCCAAACAATGG 1011
Db 666 CTGAGAGGAGAGGATACGGATGCAATCCCCCGCTCGATGTCATCAGGTCTCCAAACAATGG 725
Qy 1012 CTGTGGGCTGTTTGTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1071
Db 726 CTGTGGGCTGTTTGTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 785
Qy 1072 GAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCCCTTTAGGATCTGGCAGTGTGGTG 1131
Db 786 GAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCCCTTTAGGATCTGGCAGTGTGGTG 845
Qy 1132 GGGTCTTGGAACAACACCCATGTTCCCATGTTGGCCATGTTTCCCCAAGCAAGCTCCCT 1191
Db 846 GGGTCTTGGAACAACACCCATGTTCCCATGTTGGCCATGTTTCCCCAAGCAAGCTCCCT 905
Qy 1192 ACTCCGCAACAAGGCTCTGGCCAAACAGTGTGTCAGCTGAAGATGATGATGATGATGAT 1251
Db 906 ACTCCGCAACAAGGCTCTGGCCAAACAGTGTGTCAGCTGAAGATGATGATGATGATGAT 965
Qy 1252 TTAAGAGCTTACTTACCATCGCAACCCCGCTGGCCGCTCGAACCTTTTGGGGATGTA 1311
Db 966 TTAAGAGCTTACTTACCATCGCAACCCCGCTGGCCGCTCGAACCTTTTGGGGATGTA 1025
Qy 1312 CAGAGAGGAAGCAGCTCCGGGACAGCTCCAGTGTAAAGACTTCAAGTGGTTCCTTGGAGA 1371
Db 1026 CAGAGAGGAAGCAGCTCCGGGACAGCTCCAGTGTAAAGACTTCAAGTGGTTCCTTGGAGA 1085
Qy 1372 CTGTGTATCCAGAACTGCAATGTCCTGAGGACAGGCTTGGCTTCTTGGGATGCTCCAGA 1431
Db 1086 CTGTGTATCCAGAACTGCAATGTCCTGAGGACAGGCTTGGCTTCTTGGGATGCTCCAGA 1145
Qy 1432 ACNAGGAGTACAGACTACTGCTTGGATGATTAACCTCCCGATGAAACACAGATTTGGG 1491
Db 1146 ACNAGGAGTACAGACTACTGCTTGGATGATTAACCTCCCGATGAAACACAGATTTGGG 1205
Qy 1492 GACACAGGTCAATCTGTACCTCTCTCATGGATGGGCGAGAACTCA 1537
Db 1206 GACACAGGTCAATCTGTACCTCTCTCATGGATGGGCGAGAACTCA 1251

RESULT 7

US-10-106-698-598

; Sequence 598, Application US/10106698

; Publication No. US20030109690A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide

; FILE REFERENCE: PA005P1

; CURRENT APPLICATION NUMBER: US/10106,698

; CURRENT FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: PCT/US00/26524

; PRIOR FILING DATE: 2000-09-28

; PRIOR APPLICATION NUMBER: US 60/157,137

; PRIOR FILING DATE: 1999-09-29

; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 598
; LENGTH: 1517
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (144)..(144)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-598

Query Match 35.5%; Score 1012.2; DB 14; Length 1517;
Best Local Similarity 99.3%; Pred. No. 2.7e-224;
Matches 1008; Conservative 7; Mismatches 0; Indels 0; Gaps 0;
Qy 1115 GATCTGGCAGTGTGGGGTCTTCCGCAACAAGGCTCTGGCCAAACAGTGTTCGTCAGCTGA 1234
Db 479 GATCTGGCAGTGTGGGGTCTTCCGCAACAAGGCTCTGGCCAAACAGTGTTCGTCAGCTGA 538
Qy 1175 CCCCAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAACAGTGTTCGTCAGCTGA 1234
Db 539 CCCCAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAACAGTGTTCGTCAGCTGA 598
Qy 1235 AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1294
Db 599 AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 658
Qy 1295 ACCTTTTGGGGATGTGACAGAGGAGAGCAGTCTCCGCGAACAAGCTCCAGTGTAAAGACTT 1354
Db 659 ACCTTTTGGGGATGTGACAGAGGAGAGCAGTCTCCGCGAACAAGCTCCAGTGTAAAGACTT 718
Qy 1355 CAGTGTGTTCTGGAGACTGTGTATCCAGAACTGTCATGTCCTGCTGAGGACAGGCTTGGCTT 1414
Db 719 CAGTGTGTTCTGGAGACTGTGTATCCAGAACTGTCATGTCCTGCTGAGGACAGGCTTGGCTT 778
Qy 1415 CTTTCGGGATGCTCCAGAAACAAGGACTTAACAGACTACTGCTTTTGACTATAACCTCCCGA 1474
Db 779 CTTTCGGGATGCTCCAGAAACAAGGACTTAACAGACTACTGCTTTTGACTATAACCTCCCGA 838
Qy 1475 TGAACAACAGATTTGTGGGACACACAGGTCATTTCTGTACTCTGTCTATGGGATGGGCCAGAA 1534
Db 839 TGAACAACAGATTTGTGGGACACACAGGTCATTTCTGTACTCTGTCTATGGGATGGGCCAGAA 898
Qy 1535 TCAGTTTTTTCAGTATACGTCCTCCAGAAAGAATAACGCTATAACACCCACAGCCTGAGGG 1594
Db 899 TCAGTTTTTTCAGTATACGTCCTCCAGAAAGAATAACGCTATAACACCCACAGCCTGAGGG 958
Qy 1595 CTGCATTGCTGTGGAAGCAGGAAATGGATACCTTTATCATGCTATCTCTGCGAAGAACTGC 1654
Db 959 CTGCATTGCTGTGGAAGCAGGAAATGGATACCTTTATCATGCTATCTCTGCGAAGAACTGC 1018
Qy 1655 CCGCAGAAATCAGAAATTCATTTTGCAGGAGGATGGAATCTTTATTTTCAACGAACTGCTCA 1714
Db 1019 CCGCAGAAATCAGAAATTCATTTTGCAGGAGGATGGAATCTTTATTTTCAACGAACTGCTCA 1078
Qy 1715 GAAATGTGCTCAGGCTCCGAGGAGGAGTCCGATGAGTGTTCGTTCCACTCTTACAGAGA 1774
Db 1079 GAAATGTGCTCAGGCTCCGAGGAGGAGTCCGATGAGTGTTCGTTCCACTCTTACAGAGA 1138
Qy 1775 CTGCACCAACTCCGATCATCAGAAATGTTCTTCAAGAGGCGATGTTATGAAGCCTCGT 1834
Db 1139 CTGCACCAACTCCGATCATCAGAAATGTTCTTCAAGAGGCGATGTTATGAAGCCTCGT 1198
Qy 1835 GTATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1894
Db 1199 GTATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1258
Qy 1895 CTAAGCAGTACCCAGAAACCCCAAACTAGGCTGATGCTTTTGAAGGCGATCAAT 1954
Db 1259 CTAAGCAGTACCCAGAAACCCCAAACTAGGCTGATGCTTTTGAAGGCGATCAAT 1318

Qy 1955 TTGCATTTCTGAAAGTTGTTGGATTAGTAAATAATGGAATAAGCTTTGTACTTATT 2014
Db 1319 TTGCCATTTGTGAAGTTGTTGGATTAGTAAATAATGGAATAAGCTTTGTACTTATT 1378
Qy 2015 TTGAGAACTTTTAAATGTTTCCAAATAACCTTATTTTCAAAGGGTAAATCGTAAGATGTTA 2074
Db 1379 TTGAGAACTTTTAAATGTTTCCAAATAACCTTATTTTCAAAGGGTAAATCGTAAGATGTTA 1438
Qy 2075 ACCCTTGGTATTAGAAAAATTAACCTTATTAATATTTTCTATCAARARWAWA 2129
Db 1439 ACCCTTGGTATTAGAAAAATTAACCTTATTAATATTTTCTAWMAAAAAA 1493

RESULT 8
US-09-895-298-44
; Sequence 44, Application US/09895298
; Publication No. US20030078405A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 47 Human Secreted Proteins
; FILE REFERENCE: P2035P1
; CURRENT APPLICATION NUMBER: US/09/895,298
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/591,16
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: PCT/US99/29950
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: 60/113,006
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 60/112,809
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 1517
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (144)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-895-298-44

Query Match 35.5%; Score 1011.8; DB 11; Length 1517;
Best Local Similarity 99.2%; Pred. No. 3.3e-224;
Matches 1007; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy 1115 GATCTGGCAGTGTGGTGGGTTCTGGAAACACACCCATGTTCCCATGTTGGCCATGTTTT 1174
Db 479 GATCTGGCAGTGTGGTGGGKTCTGGAACACACCCATGTTCCCATGTTGGCCATGTTTT 538
Qy 1175 CCCAAGCAGCTCCCTACTCCGCAACAGGCTCTGGCAACAGTGTTCGTGAGCTGA 1234
Db 539 CCCAAGCAGCTCCCTACTCCGCAACAGGCTCTGGCAACAGTGTTCGTGAGCTGA 598
Qy 1235 AGTATGGATGGATGAATTTAAAGAGCTCTACTACATCGCAACCCCGTCCCGCTTGA 1294
Db 599 AGTATGGATGGATGAATTTAAAGAGCTCTACTACATCGCAACCCCGTCCCGCTTGA 658
Qy 1295 ACCTTTGGGATGTGACAGAGGAAGAGCTCCGGGAACAAGTCCAGTGAAGAATTT 1354
Db 659 ACCTTTGGGATGTGACAGAGGAAGAGCTCCGGGAACAAGTCCAGTGAAGAATTT 718
Qy 1355 CAAGTGGTCTTCGAGACTGTGATCCAGAACTGATGCTGCTGAGACAGGCTGCTT 1414
Db 719 CAAGTGGTCTTCGAGACTGTGATCCAGAACTGATGCTGCTGAGACAGGCTGCTT 778
Qy 1415 CTTTCGGATGCTCCAGAACAAAGGACTTAACAGACTACTGCTTTGACTATACCTCCCGA 1474
Db 779 CTTTCGGATGCTCCAGAACAAAGGACTTAACAGACTACTGCTTTGACTATACCTCCCGA 838
Qy 1475 TGAACACCAAGTTGTGGGACACCAAGGTCATTTCTGTACCTCTGTCAATGGGATGGGCCAGAA 1534

Db 839 TGAACACCAAGTTGTGGGACACCAAGGTCATTTCTGTACCTCTGTATGGGATGGGCCAGAA 898
Qy 1535 TCAGTTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATAACACCCACAGCCTGAGGG 1594
Db 899 TCAGTTTTTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATAACACCCACAGCCTGAGGG 958
Qy 1595 CTGCATTTCTGTGGAGCAGGAATGGAATACCTTTATCATGCTCTCTGCGAAGAACTGC 1654
Db 959 CTGCATTTCTGTGGAGCAGGAATGGAATACCTTTATCATGCTCTCTGCGAAGAACTGC 1018
Qy 1655 CCAGAGAAATCAGAAGTTTCATCTTGCAGGAGGATGGATCTTTATTTTCCACCAAGCTCAA 1714
Db 1019 CCAGAGAAATCAGAAGTTTCATCTTGCAGGAGGATGGATCTTTATTTTCCACCAAGCTCAA 1078
Qy 1715 GAAATGTGTCCAGGCTCCGAGGAAGGAGTGCAGTGCAGTTCCTGTTTCCATCTTTACGAGA 1774
Db 1079 GAAATGTGTCCAGGCTCCGAGGAAGGAGTGCAGTGCAGTTCCTGTTTCCATCTTTACGAGA 1138
Qy 1775 CTGCACAACTCGGATCATCAGAAATGGTTTCTTCAAAGAGCGCATGTTATGAAGCCTCGT 1834
Db 1139 CTGCACAACTCGGATCATCAGAAATGGTTTCTTCAAAGAGCGCATGTTATGAAGCCTCGT 1198
Qy 1835 GTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCGAGACTCTGCCCAACAAAGACTTAG 1894
Db 1199 GTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCGAGACTCTGCCCAACAAAGACTTAG 1258
Qy 1895 CTAAGCAGTGCACAGAACCCACCAAACTAGGCTGCTGATTGTTTGAAGAGGCAATCAT 1954
Db 1259 CTAAGCAGTGCACAGAACCCACCAAACTAGGCTGCTGATTGTTTGAAGAGGCAATCAT 1318
Qy 1955 TTGCCATTTGTGAAAGTTGTTGGATTTAGTAAAAATGTGAATAAGCTTTGTACTTATT 2014
Db 1319 TTGCCATTTGTGAAAGTTGTTGGATTTAGTAAAAATGTGAATAAGCTTTGTACTTATT 1378
Qy 2015 TTGAGAACTTTTAAATGTTTCCAAATAACCTTATTTTCAAAGGGTAAATCGTAAGATGTTA 2074
Db 1379 TTGAGAACTTTTAAATGTTTCCAAATAACCTTATTTTCAAAGGGTAAATCGTAAGATGTTA 1438
Qy 2075 ACCCTTGGTATTAGAAAAATTAACCTTATTAATATTTTCTATCAARARWAWA 2129
Db 1439 ACCCTTGGTATTAGAAAAATTAACCTTATTAATATTTTCTAWMAAAAAA 1493

RESULT 9
US-10-198-846-7621
; Sequence 7621, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7621
; LENGTH: 818
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2_3_16, 529, 689, 691, 709, 713, 720, 726, 739, 741, 742,
; LOCATION: 746, 749, 755, 759, 760, 761, 767, 768, 774, 775, 777, 783,
; LOCATION: 790, 791, 797, 798, 800, 801, 802, 803, 804, 805, 806,
; LOCATION: 811, 814, 815, 816, 817

122	TGCTCCAGAACAAAGGACTAACAGACTACTGCTTTTGACTATAACCCCTCCCGATGATAAAC	181
1483	AGATTGGGACACACAGGTCAATTCTGTACTCTCTGTCTATGGGATGGGCGACGAATCAGTTTT	1542
182	AGATTGGGACACACAGGTCAATTCTGTACTCTCTGTCTATGGGATGGGCGACGAATCAGTTTT	241
1543	TCGAGTACAGCTCCCGAAGAAATACGCTATTAACACCCACAGCCTGAGGGCTGCATTG	1602
242	TCGAGTACAGCTCCCGAAGAAATACGCTATTAACACCCACAGCCTGAGGGCTGCATTG	301
1603	CTGTGGAAGCAGGAATGGATACCTTATCATGCACTCTCTCGAAGAAATGCCCCAGAGA	1662
302	CTGTGGAAGCAGGAATGGATACCTTATCATGCACTCTCTCGAAGAAATGCCCCAGAGA	361
1663	ATCAGAAGTTCACTCTTCGAGGAGATGGATCTTTATTTACGAACAGCTCCAAAGAAATGTG	1722
362	ATCAGAAGTTCACTCTTCGAGGAGATGGATCTTTATTTACGAACAGCTCCAAAGAAATGTG	421
1723	TCCAGGCTCGAGGAAGAGTTCGAGTGACAGTTTCGTTTCCACTCTTACGAGACTGCACCA	1782
422	TCCAGGCTCGAGGAAGAGTTCGAGTGACAGTTTCGTTTCCACTCTTACGAGACTGCACCA	481
1783	ACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA	1826
482	ACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA	525

RESULT 13

```

US-09-777-564-172
; Sequence 172, Application US/09777564
; Patent No. US2002022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 172
; LENGTH: 632
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(632)
; OTHER INFORMATION: n = A, T, C or G
US-09-777-564-172

```

Query Match	17.4%	Score 496.4;	DB 9;	Length 632;
Best Local Similarity	90.6%;	Pred. No. 8.4e-105;		
Matches 567;	Conservative 0;	Mismatches 53;	Indels 6;	Gaps 4;
Qy	1407	CCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATAAC	1466	
Db	1	CCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATAAC	60	
Qy	1467	CTTCCCGATGAAACCGAGATTGTGGGACACGAGGTCAATTCTGTACTCTCTCATGGGATG	1526	
Db	61	CTTCCCGATGAAACCGAGATTGTGGGACACGAGGNCATTTCTGTACCTCTCTCATGGGATG	120	
Qy	1527	GGCCAGAAATC-AGTTTTTCGAGTACACGTCCCGAAGAAATACGCTATATAACCCACCA	1585	
Db	121	GGCCAGAAATCAAGTTTTTCGAGTACATTTCCCGAAGAAATACGCTATATAACCCACCA	180	
Qy	1586	GCCTGAGGGCTGCATTGCTCTGGAAAGCAGAAATGGATACCCCTTATCATGCTCTGCGA	1645	
Db	181	GCCTGAGGGCTGCATTGCTCTGGAAAGCAGAAATGGATACCCCTTACCATGCTCTGCGA	240	
Qy	1646	AGAACTGCCCCAGAGAAATCAGAAGTTTCATCTTTGCAGGAGGATGGATCTTTATTTCACGA	1705	

Db 241 AGAACTGCCCCAGAGAAATCAGAAGTTTCATCTTGCAGGAGGATGGATCTTTATTTTCACGA 300
Qy 1706 ACAGTCCAGAGAAATGTGTCCAGGCTCGAGGAAGAGTCCAGTGCAGATTTTCGTTCCACT 1765
Db 301 ACAGNCCAAGAAATGTGTCCAGGCTCGAGGAACGAGNCNAGTGCACAGTTTCGTTCCACT 360
Qy 1766 CTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATG 1825
Db 361 CTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATG 420
Qy 1826 AAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACA 1885
Db 421 AAGCCTCGTGTATCAANGAGCCCATCGAAGGAGACTGTGGAGCCAGGA-TCTGCCCAACA 479
Qy 1886 AAGACTTAGCTAAGCAGTGCACGAGACCCACCAAAACTAGGCT--GCATTTGCTTTGAAG 1943
Db 480 AAGACTTNTCTAACAGNAGCAGAAACCCACCAAAACTANGGTTGTATTNCTTTTGAAG 539
Qy 1944 AGGCAATCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTT 2001
Db 540 AAGCAATCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTT 599
Qy 2002 CTTTGTACTTATTGTGAGAACTTTT 2027
Db 600 CTTTGGAGCTTTTGTGAAACTTT 625

RESULT 14

US-09-777-564-1423

; Sequence 1423, Application US/0977564

; Patent No. US2002022591A1

; GENERAL INFORMATION:

; APPLICANT: Algate, Paul A.

; APPLICANT: Mannion, Jane

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.493

; CURRENT APPLICATION NUMBER: US/09/777,564

; CURRENT FILING DATE: 2001-02-05

; NUMBER OF SEQ ID NOS: 1730

; SOFTWARE: FastSeq for Window Version 4.0

; SEQ ID NO 1423

; LENGTH: 632

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(632)

; OTHER INFORMATION: n = A,T,C or G

US-09-777-564-1423

Query Match 17.4%; Score 496.4; DB 9; Length 632;
Best Local Similarity 90.6%; Pred. No. 8.4e-105;
Matches 567; Conservative 0; Mismatches 53; Indels 6; Gaps 4;
Qy 1407 CTTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATAAC 1466
Db 1 CTTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATAAC 60
Qy 1467 CTTCCCGATGAAACCCAGATTTGGGACACAGGTCATTCTGTACCTCTGTATGGGATG 1526
Db 61 CTTCCCGATGAAACCCAGATTTGGGACACAGGTCATTCTGTACCTCTGTATGGGATG 120
Qy 1527 GGCCAGAAATC-AGTTTTTCGAGTACACGTCGAGGAGGATGATACCCCTTATCATGCTCTGCGA 1585
Db 121 GGCCAGAAATC-AGTTTTTCGAGTACACGTCGAGGAGGATGATACCCCTTATCATGCTCTGCGA 180
Qy 1586 GCCTGAGGGCTGCATTTCTGTGGAGGAGGATGATACCCCTTATCATGCTCTGCGA 1645
Db 181 GCCTGAGGGCTGCATTTCTGTGGAGGAGGATGATACCCCTTATCATGCTCTGCGA 240
Qy 1646 AGAACTGCCCCAGAGAAATCAGAAGTTTCATCTTGCAGGAGGATGGATCTTTATTTTCACGA 1705

Db 241 AGAACTGCCCCAGAGAAATCAGAAGTTTCATCTTGCAGGAGGATGGATCTTTATTTTCACGA 300
Qy 1706 ACAGTCCAGAGAAATGTGTCCAGGCTCGAGGAAGAGTCCAGTGCAGATTTTCGTTCCACT 1765
Db 301 ACAGNCCAAGAAATGTGTCCAGGCTCGAGGAACGAGNCNAGTGCACAGTTTCGTTCCACT 360
Qy 1766 CTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATG 1825
Db 361 CTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATG 420
Qy 1826 AAGCCTCGTGTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCAGGACTCTGCCCAACA 1885
Db 421 AAGCCTCGTGTATCAANGAGCCCATCGAAGGAGACTGTGGAGCCAGGA-TCTGCCCAACA 479
Qy 1886 AAGACTTAGCTAAGCAGTGCACGAGACCCACCAAAACTAGGCT--GCATTTGCTTTGAAG 1943
Db 480 AAGACTTNTCTAACAGNAGCAGAAACCCACCAAAACTANGGTTGTATTNCTTTTGAAG 539
Qy 1944 AGGCAATCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTT 2001
Db 540 AAGCAATCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTTTGCATTT 599
Qy 2002 CTTTGTACTTATTGTGAGAACTTTT 2027
Db 600 CTTTGGAGCTTTTGTGAAACTTT 625

RESULT 15

US-10-015-219-172

; Sequence 172, Application US/10015219

; Publication No. US20030008299A1

; GENERAL INFORMATION:

; APPLICANT: Algate, Paul A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.493C1

; CURRENT APPLICATION NUMBER: US/10/015,219

; CURRENT FILING DATE: 2002-03-02

; NUMBER OF SEQ ID NOS: 1739

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 172

; LENGTH: 632

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: 95, 305, 338, 340, 437, 487, 496, 513, 520, 530, 552, 604

; OTHER INFORMATION: n = A,T,C or G

US-10-015-219-172

Query Match 17.4%; Score 496.4; DB 14; Length 632;
Best Local Similarity 90.6%; Pred. No. 8.4e-105;
Matches 567; Conservative 0; Mismatches 53; Indels 6; Gaps 4;
Qy 1407 CTTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATAAC 1466
Db 1 CTTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATAAC 60
Qy 1467 CTTCCCGATGAAACCCAGATTTGGGACACAGGTCATTCTGTACCTCTGTATGGGATG 1526
Db 61 CTTCCCGATGAAACCCAGATTTGGGACACAGGTCATTCTGTACCTCTGTATGGGATG 120
Qy 1527 GGCCAGAAATC-AGTTTTTCGAGTACACGTCGAGGAGGATGATACCCCTTATCATGCTCTGCGA 1585
Db 121 GGCCAGAAATC-AGTTTTTCGAGTACACGTCGAGGAGGATGATACCCCTTATCATGCTCTGCGA 180
Qy 1586 GCCTGAGGGCTGCATTTCTGTGGAGGAGGATGATACCCCTTATCATGCTCTGCGA 1645
Db 181 GCCTGAGGGCTGCATTTCTGTGGAGGAGGATGATACCCCTTATCATGCTCTGCGA 240
Qy 1646 AGAACTGCCCCAGAGAAATCAGAAGTTTCATCTTGCAGGAGGATGGATCTTTATTTTCACGA 1705

Db 241 AGAACTGCCCCAGAGATCAGAAGTTTCATCTTCAGAGGATGGATCTTTATTTACGA 300
Qy 1706 ACAGTCCAGAAATGTGTCCAGGCTGCGAGAGAGAGTCGAGTCACAGTTTTTCGTTCCACT 1765
Db 301 ACAGNCCAAGAAATGTGTCCAGGCTGCGAGAGAGAGNAGTCACAGTTTTTCGTTCCACT 360
Qy 1766 CTTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATGTTATG 1825
Db 361 CTTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATGTTATG 420
Qy 1826 AAGCTCTGTGTATCAAGGAGCCCATCGAAGGAGACTGTGGAGCCGAGACTCTGCCCAACA 1885
Db 421 AAGCTCTGTGTATCAANGAGCCCATCGAAGGAGACTGTGGAGCCGAGACTCTGCCCAACA 479
Qy 1886 AAGACTTAGCTAAGCAGTGCACGAGCCCAACCAAAAACTAGGCT--GCATTTGCTTTGAAG 1943
Db 480 AAGACTTNCCTAACCAAGNGACCAGAAACCCACCAAACTANGTTGTATTNCTTTTGAAG 539
Qy 1944 AGGCAATCATTTTCCCATTTGTGAAAGTTGTGTTGGATTAGT--AAAAATGTGAATAAG 2001
Db 540 AAGCAATCATTTNGCCTTTTGTGAAAGTGTGTTGGATTAAATTAATAAGGGGAATAAA 599
Qy 2002 CTTTGTACTTATTTTGAGAACTTTTT 2027
Db 600 CTTTNGGACTTTTTTTTGAAAACTTT 625

Search completed: December 8, 2003, 01:56:02
Job time : 873.526 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 12:46:13 ; Search time 5923.86 Seconds
(without alignments)
11693.015 Million cell updates/sec

Title: US-10-074-527-1

Perfect score: 2850

Sequence: 1 ccggctgggtaccactataa.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1303.8	45.7	2217	11 AK033638	AK033638 Mus muscu
2	1297.8	45.5	2074	11 AK042133	AK042133 Mus muscu
3	1257	44.1	2034	11 AK082014	AK082014 Mus muscu
4	1061	37.2	1781	11 BC025639	BC025639 Mus muscu

ALIGNMENTS

RESULT 1

AK033638

LOCUS

DEFINITION

AK033638

Mus musculus adult male cecum cDNA, RIKEN full-length enriched library, clone:9130206E10 product:hypothetical Glycosyl transferase, family 2 containing protein, full insert sequence.

ACCESSION

AK033638.1

VERSION

AK033638.1

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

1

Carninci,P. and Hayashizaki,Y.

AUTHORS

TITLE

High-efficiency full-length cDNA cloning

JOURNAL

Meth. Enzymol. 303, 19-44 (1999)

MEDLINE

92279253

REFERENCE

2

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

AUTHORS

TITLE

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL

Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE

20499374

PUBMED

11042159

5	992	34.8	2064	11	BC024475
6	943.8	33.1	1752	11	AK043977
7	868	30.5	1909	14	CD251293
8	738.2	25.9	745	12	CB243715
9	723	25.4	745	12	BM978800
10	718.6	25.2	782	13	BU854444
11	705.4	24.8	711	12	BM977212
12	699.8	24.6	705	12	BM970195
13	666.6	23.4	741	10	AW69574
14	660.8	23.2	676	10	BE677813
15	646	22.7	1201	13	BU343372
16	626	22.0	932	10	EG497683
17	610.8	21.4	967	10	EG167520
18	599.4	21.0	868	10	BE867781
19	580	20.4	1201	13	BU343371
20	577.4	20.3	792	10	EG687328
21	568.6	20.0	656	10	EG499007
22	566.4	19.9	692	14	CB165548
23	557	19.5	559	10	BE218607
24	548.2	19.2	561	9	AI800923
25	541.6	19.0	770	14	CD299534
26	538.6	18.9	708	12	EG777267
27	530.8	18.6	790	10	BF102783
28	524.8	18.4	536	10	BE673969
29	520.8	18.3	2607	11	AK033494
30	520.8	18.3	2607	11	AK033494
31	519.8	18.2	526	9	AI690020
32	514	18.0	522	9	AI566491
33	508	17.8	661	12	BI154821
34	505	17.7	518	10	BE502927
35	503.2	17.7	721	12	EG868185
36	491.8	17.3	495	10	BF197099
37	482	16.9	818	10	BF246502
38	480.6	16.9	511	9	AI863865
39	479	16.8	690	10	BF031122
40	472.4	16.6	477	9	AW612896
41	465.8	16.3	549	13	BU679643
42	457	16.0	457	10	BF222363
43	456.4	16.0	506	10	BE940623
44	456.2	16.0	461	9	AI223049
45	450.8	15.8	496	9	AA429394

AK033638 2217 bp mRNA linear HTC 05-DEC-2002
Mus musculus adult male cecum cDNA, RIKEN full-length enriched library, clone:9130206E10 product:hypothetical Glycosyl transferase, family 2 containing protein, full insert sequence.

AK033638 GI:26329324

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1

Carninci,P. and Hayashizaki,Y.

AUTHORS

TITLE

High-efficiency full-length cDNA cloning

JOURNAL

Meth. Enzymol. 303, 19-44 (1999)

MEDLINE

92279253

REFERENCE

2

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

AUTHORS

TITLE

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL

Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE

20499374

PUBMED

11042159

REFERENCE
AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

TITLE

JOURNAL
MEDLINEPUBMED
REFERENCE
AUTHORS

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Tomita, M.,
Quackenbush, J., Schrim, L.M., Staub, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
Marichonni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schombach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyo-oka, K., Wang, K.H., Weitz, C., Wittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.,
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851

TITLE

JOURNAL
MEDLINEPUBMED
REFERENCE
AUTHORS

5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2217)

JOURNAL
MEDLINEPUBMED
REFERENCE
AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hangaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayashi, N., Hiramoto, K., Hiraoka, I., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.
Direct Submission

TITLE

JOURNAL

MEDLINE

PUBMED
COMMENT

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216]
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers

FEATURES

source

1. 2217
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:9130206E10"
/db_xref="taxon:10090"
/clone="9130206E10"
/sex="male"
/tissue_type="cesum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
22. 1752

CDS

/notes="unnamed protein product; hypothetical Glycosyl
transferase, family 2 containing protein
(InterPro|IPR001173, evidence: InterPro)
putative"

/codon_start=1
/protein_id="BAC28401.1"
/db_xref="GI:26329325"

/translations="MNGRAVRRRCPLRRRGLALLALAGLALLRARSRSCTVD
PGPRPTPLGRHEPVLPRPLPADALGAHEAVRLQQLQELRLQESVKQHOINYL
SDRISLHRLPGRWNLPCREVKYDNLPKTSVVIAYFANAMSTLLRTVSVLETSPL
ILLEEVILVDYSDRLKRLNELSQLPRLIRASREGLVRARLLGASARGV
LTFDCHCEHEGLPLQRIHEKESAVVCPVIDWNTFYLGNSGPGQGGFDM
RLVFTWHVQPERQSMRSPIDVIRSTWAGLVFAVSKRYDFYLSGDTGMEVGGFN
LRFSPWOCGGTLETHPCSHVGHVFPKQAPYSKALANSVAEVMDFEKLYYLH
RNPQARLEPGDVTERRKRLAKLQCKDFKFLDTVYPLHVPEDRPGFGLNQRGL
CYCLDYNPNENHVEGHVLLYLCHGQNGQFPEYTRKIRVINTROPEACITVEDGK
DTUMLDCKRETPENQEFILQEDGTLVHRSRCKATEKVLQNGFAPYLRDCTNSN
QRWFKERMS"
2198. 2203
/notes="putative"
2217

polyA_signal

polyA_site

BASE COUNT 503 a 543 c 666 g 505 t

ORIGIN

Query Match 45.7%; Score 1303.8; DB 11; Length 2217;
Best Local Similarity 76.1%; Pred. No. 6.7e-133;
Matches 1771; Conservative 4; Mismatches 431; Indels 121; Gaps 9;
Qy 62 GGCTGCACTTTGGCGGCGCATGTGGGGGCGCACGGCGCGCGCGTCCCGCGGGAAC 121
Db 3 GGGTGCAGCTAGCGGGTGCATGTGGGGCGTGGCGTGAGGAGCGCTCCCGCGGGCG 62
Qy 122 GCGCGCGCGCGGCGCGCGCTGTGGTCTCTTGGCGCTACTGGCGTGGCGGGCTGGG 181
Db 63 GCGCGCGCGCGGAGCGCTTG-----CTGGCACTCTGGCCCTGGCGAGGGCTGGG 113
Qy 182 CTGGTGTCTCGGCGCGAGCGTGGGGCGCGGGGCTGCCGAGCCGGACCCCGCGCG 241
Db 114 CGCGTGTCTCGGCG-----TCGACGAGTGGGGAACGGTGACCCCGGACCCCGCG 167
Qy 242 CACCCCGCGCGCGCGCGCGCGCGTTCATCCCGCGCGCGCGTGGCGGCGGACGC 301
Db 168 GACTCTCTCCCGCGCGTCCAGCGCGAGTGTGCCAGCGCGCGCTGCCCGCGGATGC 227
Qy 302 GCTGGCGCGCGCGCGCGCGCGTGGCTGCAGCTGCAGGCGGAGAGCTGGCGGCTGCA 361
Db 228 TCTGGGAGCGCATGGGAGGCGCGTGGCGTGCAGCTGCAGGCGGAGAGCTGGCGGCTGCA 287
Qy 362 GGAGGAGAGCGTGGCGCTGCACAGATTAACATCTACTCTCAGCGACCGCATCTCACTGCA 421
Db 288 GGAGGAGAGCGTGAAGCAGCAGCCAGATCAACATCTATCTGAGCGACCGTATCTCGCTGCA 347
Qy 422 CCGCGCGCTCCCGAGCGCTGGAAACCGCTGTGTCAAGAGAGAAATATGATTATGATAA 481
Db 348 CCGCGGACTCCCGAGCGCTGGAAACCTCTGTGTGAGAGAGGTGAAATACGATTATGATAA 407
Qy 482 TTGCGCGGAGCAGCATCTGTTATCATAGCATTTTATTAATGAAGCCCTGGTCAACTCTCTTCG 541
Db 408 CCTGCGGCAAGCACTTCTGTGTGATAGCATTTTATTAATGAAGCCCTGGTCCACACTCTCTTCG 467

QY 542 GACAGTTTACAGTCTCTTGAGACATCCCGGATATCTCTGTAGAGAGTATCTTGT 601
Db 468 GACAGTTTACAGTCTCTTGAGACATCCCGGATATCTCTGTAGAGAGTATCTTGT 527
QY 602 AGATGACTACAGTATAGAGACACCTTGAAGAGCGCTTGGCCAAATGAGCTTTCGGGACT 661
Db 528 AGATGACTACAGTATAGAGACACCTTGAAGAGCGCTTGGCCAAATGAGCTTTCAGACT 587
QY 662 GCCCAAGGTGCGCTGATCCCGGCCAAACAGAGAGAGGGCCCTGTTGGAGCCCGGCTGCT 721
Db 588 CCCCAAGGTGCGCTGATCCCGGCCAAACAGAGAGAGGGCCCTGTTGGAGCCCGGCTGCT 647
QY 722 GGGGGCTCTCGCGGAGGGGCGATGTTCTGACCTTCTGACACTGCTGAGTGGCA 781
Db 648 GGGAGCTCTCGCGGAGGGGCGATGTTCTGACCTTCTGAGCTGCTGAGTGGCA 707
QY 782 CGAAGGGTGGCTGGAGCGCTGCTGCGAGAGATCCATGAGAGGAGTGGGAGTGGT 841
Db 708 TGAGGGTGGCTGGAGCGCTGCTGCGAGAGATCCATGAGAGGAGTGGGAGTGGT 767
QY 842 CCCGGTGAATGATGATCGATCGAATCGAATCGAATCGAATCGAATCGAATCGAAT 901
Db 768 CCCGGTGAATGATGATCGAATCGAATCGAATCGAATCGAATCGAATCGAAT 827
QY 902 CCAGATCGGCGGTTTCCAGTGGAGGCTGGTGTTCAGCTGGCACAGTTCCTGAGAGGA 961
Db 828 GCAGATGGGCGGTTTCCAGTGGAGGCTGGTGTTCAGCTGGCACAGTTCCTGAGAGGA 887
QY 962 GAGGATACGGATCAATCCCGGCTGATGTCATGAGTCTCCAAATGCGTGGGCT 1021
Db 888 GCGGAGTGGTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 947
QY 1022 GTTGTCTGTGAGTGAAGAAATATTTTGAATATCTGGGCTCTTATGATACAGGAATGGAAGT 1081
Db 948 GTTGTCTGTGAGTGAAGAAATATTTTGAATATCTGGGCTCTTATGATACAGGAATGGAAGT 1007
QY 1082 TTGGGGAGGAGAAACCTCGAATTTTCTGATGATGATGATGATGATGATGATGAT 1141
Db 1008 CTGGGGAGGAGAAACCTCGAATTTTCTGATGATGATGATGATGATGATGATGAT 1067
QY 1142 AACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCGCCAA 1201
Db 1068 AACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCGCCAA 1127
QY 1202 CAAGGCTCTGCGCCAAACAGTGTTCGTCAGCTGGAAGTATGATGATGATGATGATGAT 1261
Db 1128 CAAGGCTCTGCGCCAAACAGTGTTCGTCAGCTGGAAGTATGATGATGATGATGATGAT 1187
QY 1262 CTACTACCTCGCAACCCCGTGGCCGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 1321
Db 1188 CTACTACCTCGCAACCCCGTGGCCGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 1247
QY 1322 GCAGCTCCGGGACAGCTCCAGTGTGAAGCTTCAAGTGTGTTCTTGAGAGCTGTGTATCC 1381
Db 1248 GAAGCTTCCGGCTTAAGCTCCAGTGTGAAGCTTCAAGTGTGTTCTTGAGAGCTGTGTATCC 1307
QY 1382 AGAATCGATGTGCTCGAGGACAGGCTGCTTCTCGGATGCTCCAGAACAAAGACT 1441
Db 1308 AGAATCGATGTGCTCGAGGACAGGCTGCTTCTCGGATGCTCCAGAACAAAGACT 1367
QY 1442 AACAGACTACTGCTTTGATATATCCCTCCCGATGAAACCAAGATTTGGGACACCAAGT 1501
Db 1368 AAGAGGCTACTGCTTCTGACTACATCTCTCCCAATGAAACCAAGATTTGGGACACCAAGT 1427
QY 1502 CATTCCTACTCTGCTATGAGTGGGACAGTCTTCTGAGTACAGTCTCCAGTCCAGAA 1561
Db 1428 CCTTCTACTCTGCTATGAGTGGGACAGTCTTCTGAGTACAGTCTCCAGTCCAGAA 1487
QY 1562 AGAATACGCTATTAACACCCAGGCTGAGGGCTGATGCTGCTGGAAGCAGGAATGGA 1621
Db 1488 AGAATACGCTATTAACACCCAGGCTGAGGGCTGATGCTGCTGGAAGCAGGAATGGA 1547
QY 1622 TACCCCTTATCATGATCTCTGCGAAGAACTGCCCCAGAGAAATCAGAAGTTCTCTTGA 1681

Db 1548 TACCCCTTATCATGATCTCTGCGAGAGACCGTCCAGAGAACCCAGAGTTCTATCTTACA 1607
QY 1682 GGAGGATGGATCTTTTATTTTACCAAGACAGTCAAGAAATGTGTCCAGCTGCGAGGAAGGA 1741
Db 1608 GGAGGATGGATCTTTTATTTTACCAAGACAGTCAAGAAATGTGTGGAGGCCACAGAGAGGT 1667
QY 1742 GTCGAGTGACAGTTTCTGCTTCTTACCTCTTACGAGAGTGCACCAACTCGGATCATCAGAAATG 1801
Db 1668 GTTAGACAACGGCTTTGACACCTTACGGGACTGTACCAACTCAGATACCAAGAGGTG 1727
QY 1802 GTTCTTCAAGAGGCGATGTTATGAGGCTCTGTGTATCAAGGAGGCCCATCGAAGGAGACT 1861
Db 1728 GTTCTTCAAGAGGCGATGTTATGAGGCTCTGTGTATCAAGGAGGCCCATCGAAGGAGACT 1779
QY 1862 GTGAGGCGAGGACTCTGCCCCAACAAAGACTTAGCTAAGCAGTGCACCAAGACCCACCAAAA 1921
Db 1780 TCCAGGCTGTGAGCTGTGCCCCAACAAAGACTTAGCTAAGCAGTGCACCAAGACCCACCAAAA 1805
QY 1922 ACTAGGCTGATGCTTGTGTTGAAGGCGCAATCATTTTGCATTTGTGAAAGTTGTGTTGGAT 1981
Db 1806 -----TGAGATGGACAGCGTACACGGTTGTCTTGTGAA---TGTGCTGGGT 1850
QY 1982 TTAGTAAATGTGAATAGCTTTTGTACTTATTTTGAAGACTTTTAAATGTTCCAAAT 2041
Db 1851 GTCAAGAGTGTGAATAGCTTTTGTACTTATTTTGAAGACTTTTAAATGTTCCAAAT 1910
QY 2042 ACCCTATTTTCAAGGGTAACTCGTAGATGCTTAAACCTCGTATTTTGAAGAAATTTAAACC 2101
Db 1911 ATCTGCTTTCTGAGGCAATCTTAGATGTTG---CCTTTGGTATTTGGAAGG----- 1962
QY 2102 TTATAATATTTTCTATCAARWAWATTTTACAGTCGTCGCTTTTACTCTCATTAGCA 2161
Db 1963 -----ATATTTTCTATCAAGATGTGCAATTTTACAAATGTCCTTTTATTTCTCATTAGCA 2017
QY 2162 AAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTTCCAGGTCAGAAATATCTGCA 2221
Db 2018 AAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTTCCAGGTCAGAAATATCTGCA 2057
QY 2222 TGGGTGGAATCAGGTTTCAAGCAACGTTTCTGATTTTCAAGTAACTGATTAATACCTCAGTCGG 2281
Db 2058 -----TGGGTGTTTGTAGCACATTTTACAACTGTTAATACCTCAATTTG 2106
QY 2282 GGTAAAGTTTTCCTAGTATAGAGAGACTGCTGCTAGGAACTTGTATTTGATTTT-----A 2336
Db 2107 AGTTAGATTTCTCTCAGGGGATGAGTGGACACTAGAAATACTGTTGATTTCTTCTCAGGT 2166
QY 2337 TTAGGCTTATGAGATCTTCTAGATGATTTTAAAAGAAATGCTTTT 2383
Db 2167 GTGAGTGGGTGAGATGATCTAGATTCATACATAAAGAAATGCTTTT 2213

RESULT 2

AK042133 2074 bp mRNA linear HTC 05-DEC-2002
LOCUS Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
DEFINITION library, clone:A630062B03 product:hypothetical Glycosyl
transferase, family 2 containing protein, full insert sequence.
AK042133
VERSION AK042133.1 GI:26334956
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE Carninci, P. and Hayashizaki, Y.
AUTHORS High-efficiency full-length cDNA cloning
TITLE Meth. Enzymol. 303, 19-44 (1999)
JOURNAL 92279253
MEDLINE 10349636
PUBMED
2
REFERENCE Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
AUTHORS

Qy	474	TATGATAAATTGCCCGAGACATCTGTTATCATAGCATTTTATATGAAGCCTGGTCAACT	533
Db	432	TATGATAAACCCTGCCCAAGACTTCCTGTTGTCATAGCATTTTATATGAAGCCTGGTCCACA	491
Qy	534	CTCCTTCGGACAGTTTACAGTGTCTTTGAGACATCCCGGATATCTCTGCTAGAGAAGATG	593
Db	492	CTCCTTCGGACAGTTTACAGTGTCTTTGAGACCTCCCTGACATCTCTGCTGGAGAGGTC	551
Qy	594	ATCCTTGTAAGTACTACAGTATAGAGAGCACTGAAAGGAGCGCTTGGCCAAATGAGCTT	653
Db	552	ATTCTGTTAGATGACTACAGCGACAGAGAGCACCTAAAGGAAACGCTTTGGCCAAAGAGCTG	611
Qy	654	TCGGGACTGCCCAAGGTGGCGCTGATCCGCGCCAAACAGAGAGAGGCGCTGGTCGAGCC	713
Db	612	TCACAGCTCCCCAAGGTGGCGCTGATCCGTGCTAGCAGAGAGAGGCGCTAGTGCAGACC	671
Qy	714	CGGCTGCTGGGGCGTCTGGCGGCAGGGGCGATGTTCTGACCTTCTCTGGACTGTCACTGT	773
Db	672	CGGCTACTGGAGGCTCTCGCGCCAGGGCGAAGTGTGACCTTCTTAGACTGTCACTGT	731
Qy	774	GAGTGCCAGAAAGGTGGCTGGAGCGCGTGTGCAGAGGATTCATAGAGGAGTTCGGCA	833
Db	732	GAATGTATAGAGGGGTGGCTGGAGCCTCTGCTGCAGAGGATCCACGAGAAGGAGTCCGCA	791
Qy	834	GTGTTGTGCCCGGTATTTGATGTGATCGACTGGAAACACCTTCGAAATACCTGGGAACTCC	893
Db	792	GTGTTGTGCCCTGTTATTCGATGTGATTTGACTTGGAAACACCTTTGAGTACCTTGGGCAATCG	851
Qy	894	GGGAGGCCCAGATTCGGCGGTTCGCACTGGAGGCTGGTGTTCACGTGGCAACAGTTTCCCT	953
Db	852	GGAGAGCCGCAGATTGGCGGCTTTGACTGGCGGTGGTATTTACGTGGCAGCGTGTCCC	911
Qy	954	GAGAGGAGAGGATACGGATGCAATCCCGTGCATGCATCAGGTCTCAACAAATGGCT	1013
Db	912	CAGCGGAGCGGCAGTGCATCGCGTCCCGATCGATGTATTACAGTCTCAAACTATGGCT	971
Qy	1014	GGTGGGCTGTTTGTGTGAGTAAAGAAATTTTGAATATCTGGGGTCTTATGATACAGGA	1073
Db	972	GGGGGCTGTTTGTGTGAGTAAAGAAATTTTGAATATCTGGGGTCTTACGATACAGGA	1031
Qy	1074	ATGGAAGTTTGGGAGGAGAAACCTCGAAATTTTCTTTAGGATCTGGCAGTGTGGTGGG	1133
Db	1032	ATGGAAGTCTGGGAGGAGAAACCTTTGAGTCTTCTTTAGGATCTGCAGTGTGGTGGC	1091
Qy	1134	GTTCGGAACAACCCATGTTCCCATGTTGGCCATGTTTTCCCAAGCAAGCTCCCTAC	1193
Db	1092	ACTCTGGAACGCAACCCGTCTCCATGTGGGCCACGTCTTCCCTAAGCAAGCTCCCTAT	1151
Qy	1194	TCCCGCAACAAGGCTCTGGCCAAACAGTGTTCGTGCAGCTGAAGTATCGATGGATGAATTT	1253
Db	1152	TCCCGTAGCAAGGCCCTGGCCAAACAGTGTTCGAGCTCGAAGTGTGGATGGATGAATTT	1211
Qy	1254	AAAGAGCTCTACTACCATCGCAACCCCGCTGCCGCTTGGAAACCTTTTGGGGATGTGACA	1313
Db	1212	AAAGAACTCTACTACCAACCGCAATCCCCAGGCCCGCTGGAAACCTTTTGGGACGTGACA	1271
Qy	1314	GAGAGGAGAGCTCCGGGACAGCTCCAGTGTAAAGACTTCAAGTGGTTCCTGGAGACT	1373
Db	1272	GAGAGGAAGAAGCTTCGGGCTAAGCTCCAGTGTAAAGACTTCAAGTGGTTCCTTAGATACA	1331
Qy	1374	GTGTATCCAGAACTGCATGTGCCTGAGGACAGGGCTGGCTTCTTTCGGGGATGCTCCAGAAC	1433
Db	1332	GTGTACCCAGAAATGACAGTGCACAGTGCACAGGAGACAGGGCTGGCTTCTTTGGGATGCTCCAGAAC	1391
Qy	1434	AAAGGACTAAACAGACTACTGCTTTGACTATAACCCCTCCGATGAAACCAGATTTGGGGA	1493
Db	1392	AGAGGCTAAGAGGGTACTGCCTTGACTACTCAATCCTCCCAATGAAACCACATGTTGAAGGC	1451
Qy	1494	CACAGGTCAATCTGTACCTCTGTCACTGGATGGGCCAGAAATCAGTTTTTTCGAGTACACG	1553
Db	1452	CACAGGTCTCTGTACCTCTGCCATGGGATGGGTCAAGAACCAAGTTTTTTCGAGTATACG	1511

QY	1554	TCCGAGAAAGAAATACGCTATAACACCCACCGCCTGAGGGCTGCATTCGCTGTGGAGCA	1613
Db	1512	ACCCGGAAGAAATACGCTATAACACCCGACCGCAAGCCTTCGATCAACCGTGGAGGAC	1571
QY	1614	GGAAATGATACCCCTTATCATGTCATCTCTCGGAAGAAACTGCCCCAGAGAAATCAGAAGTTC	1673
Db	1572	GGGAAGATACCTTGTTCATGGATCTCTGCCGAGAGCCGTCCAGAGAAACGAGGATTC	1631
QY	1674	ATCTTGAGGAGATGGATCTTTTATTTTCAAGAACAGTCCAAGAAATGTGTCCAGGCTGG	1733
Db	1632	ATCTCAGAGAGATGGCACTTTTAGTTTCAAGACACAGCAGGAAATGTGTGGAGGCCACA	1691
QY	1734	AGGAAGAGTCGAGTGACAGTTCGTTCCTCTTACGAGACTGCACCAACTCGGATCAT	1793
Db	1692	GAGAAGGTGTGTAGAACCGCTTTTGACCATATCTTACGGGACTGTACCAACTCAGATAAC	1751
QY	1794	CAGAAATGGTTCCTTCAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGGAGCCCATCGA	1853
Db	1752	CAGAGGTGGTTCCTTCAAGAGCGCATGTCGTAGTGATCGCCT-----TGCCAAAGA	1803
QY	1854	AGGAGACTGTGGAGCAGGAGCTCTGCCCAACAAAGACTTAGCTAAGCAGTGACCGAACCC	1913
Db	1804	AGGAGCCCTCCAGCGCTGTGGACTG-----1828	
QY	1914	CACCAAAAACCTAGGCTGCATTCGTTTGAAGAGCAATCATTTTGCCATTTGTGAAGTTG	1973
Db	1829	-----TGTCGCCAATGAGATGGACAGCGTACACGGTGTGCTTCGTGGA-----TG	1874
QY	1974	TGTTGGATTTAGTAAAAATGTGAATAAGCTTTGTACTTATTTTGGAGAACTTTTAAATGT	2033
Db	1875	TGCTGGGTGTCAACAGAGTGTGAATAAGCTTTGTACAGATTTTGAAGAACTTTAAAAATGT	1934
QY	2034	TCCAAAATACCCCTATTTTCAAAGGGTAATCGTAAGATGTTAACCCCTGGTATTATTAGAAA	2093
Db	1935	CTCCTGGCATCCTGCTTTCTGAGGGCAATCCTAAGATGTT- GCCTTTGGTATTTTGGAA--	1991
QY	2094	TTAAACCTTATAATATTTTCTATCAABAWRAWATTTTACAGTCGTGCCCTTTTACTCT	2153
Db	1992	-----AGGATATTTTCCATCAAGATGTCATTTTACATGGTGCCTTTTATCT	2041
QY	2154	CATTAGCAAAAAAGATAAAGATTTTATTTTGGT	2186
Db	2042	CATTAGCAAAAAAGGTAAATACCTTTATTTTGTAT	2074

RESULT	3
LOCUS	AK082014
DEFINITION	Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:c130094P07 product:hypothetical Glycosyl transferase, family 2 containing protein, full insert sequence.
ACCESSION	AK082014
VERSION	AK082014.1 GI:26100345
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.; Mus.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki.Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, Y., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplex capillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gajobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Asahara, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsu, Y., Nikaide, I., Pesole, G., Tonita, M., Quackenbush, J., Schmitt, L., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M., Brownstein, M., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hum, D., Kaniya, M., Lee, N., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombert, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K., Suzuki, H., Toyooka, K., Wang, K., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)
 21085660
 11217851

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 2034)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hasegawa, T., Hara, A., Hashizume, W., Hayashida, K., Hayashida, K., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@ac.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 Location/Qualifiers
 1. .2034

FEATURES

SOURCE

/organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM DB:C130094P07"
 /db_xref="taxon:10090"
 /clone="C130094P07"
 /tissue_type="head"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="16 days embryo"
 misc_feature 66..1797
 /notes="hypoetical Glycosyl transferase, family 2 containing protein (InterPro|IPR001173, evidence: InterPro)
 putative"
 BASE COUNT 439 a 528 c 635 g 432 t
 ORIGIN
 Query Match 44.1%; Score 1257; DB 11; Length 2034;
 Best Local Similarity 82.6%; Pred. No. 8,4e-128;
 Matches 1480; Conservative 0; Mismatches 295; Indels 16; Gaps 3;
 QY 54 AGATCGCTGGCTGCAGTTGGCGGGCGCATGTGGGGGGGACGCGGGCGGGCGCGCTGCCG 113
 DB 39 AGCCCGCGGGGTGCAGTACGCGGTGCATGTGGGGGGCGTGGGTGAGGAGCGCGTCCCG 98
 QY 114 CGGGAACCTGGCGCGCGCGGGGAGGCGCTGTGTGCTCTCTGGCTTCTCTGGCTTCTGGCC 173
 DB 99 CGGGGGCTGGCGCGCGCGGGGAGGCGTTG-----CTGGCACTGCTGGCGCTGGCA 149
 QY 174 GGGCTGGGCTCGTGTGTGGCGCGCAGCGTGGGGCGCGGGCGCGGGCGTGGCGAGCGCGGA 233
 DB 150 GGGCTGGCGCGCTGTCTGGCGGC-----TCGAAGCAGGTCGGGGACGTTGGACCGCGGA 203
 QY 234 CCCCCTGGCG 293
 DB 204 CCCCCTGGCGCGCTCTCTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 263
 QY 294 GCGAAGCGCTGGCG 353
 DB 264 GCGATGCTCTGGAGCGCGATGGGAGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 323
 QY 354 CGGCTGCAGGAGGAGCGTGGCGCGTGCACAGATTAACTATCACTTACCTCAGCGACCGCATC 413
 DB 324 CGGCTGCAGGAGGAGCGTGAAGCAGCAGATCACTATCTTCTGAGCGACCGTATC 383
 QY 414 TCACTGCACCG 473
 DB 384 TCGCTGCACCG 443
 QY 474 TATGATTAATTTGCCAGGACATCTGTTAT-CATAGCATTTTATTAATGAAGCCTGGTCAAC 532
 DB 444 TATGATTAATTTGCCAGGACATCTGTTGTCATAGCATTTTATTAATGAAGCCTGGTCCAC 503
 QY 533 TCTCTTTTCGACAGTTTACAGTGTCTTTGAGACATCCCGGATATCTCTGTAGAGAGAGT 592
 DB 504 ACTCTTTTCGACAGTTTACAGTGTCTTTGAGACATCCCGGATATCTCTGTAGAGAGAGT 563
 QY 593 GATCTTGTAGTACGTACAGTGTAGAGCAGCTGAAGGCGCTTGGCCATGAGCT 652
 DB 564 CATTTCTGTAGTACGTACAGCAGAGACACCTTAAGGAAGCGTTGGCCAAACAGAGCT 623
 QY 653 TTGGGAGCTCCCAAGGTGGCGCTGATCCGCGCGCAACAAGAGAGAGCGCTGGTGGCGAGC 712
 DB 624 GTACAGCTCCCAAGGTGGCGCTGATCCGCTGTAGCAGAGAGAGCGCTTAGTGGGAGC 683
 QY 713 CCGGCTGTGGGGCGCTTCGCGCGAGGGCGAGTGTCTGACCTTCTTGGACTGTCTACTG 772
 DB 684 CCGGCTACTGGAGCGCTTCGCGCGAGGGCGAGTGTCTGACCTTCTTGGACTGTCTACTG 743
 QY 773 TGAGTGCACGAGAGGTGGCTGGAGCGCGCTGTCTGAGAGGATCCATGAGAGAGGTGGCG 832
 DB 744 TGAATGTATGAGGGGTGGCTGGAGCGCTCTGCTGAGAGGATCCACGAGAGAGGTGGCG 803

Db 246 GGTAGATGACTACAGCGACAGAGACACCTAAAGAACGCTTGGCCAAACGAGCTGTCA 305
Qy 659 ACTGCCCAAGGTGGCTGATCCGCGCCAAACAGAGAGAGGGGCTGTGCGAGCCGGCT 718
Db 306 GCTCCCCAAGGTGGCTGATCCGCTGTAGCAGAGAGAGGGGCTGTGCGAGCCGGCT 365
Qy 719 GCTGGGGGCGTCTGCGCGAGGGGCGATGTTCTGACCTTCTTGACCTGTCTGACCTGTGAGTG 778
Db 366 ACTGGGAGCTCTTCCGCGCAGGGGCGAAGTGTCTGACCTTCTTAGACTGTCTGCTGTGATG 425
Qy 779 CCACGAAGGGTGGCTGGAGCGCTGTCTGCGAGAGATCCATGAAGAGAGTGGCAGTGGT 838
Db 426 TCATGAGGGTGGCTGGAGCTCTGCTGCGAGAGATCCACGAGAGGAGTCCGCGAGTCT 485
Qy 839 GTGCCCGGTGATGATGTGATCGATCGAACACCTTCGAATACCTGGGGAACTCCGGGA 898
Db 486 GTGCCCTGTTATCGATGTGATGTAAGTCTGGAACACCTTTGAGTACCTGGGCAACTCGGGA 545
Qy 899 GCGCCAGATGGCGGTTTTCGACTGGAGCTGGTGTTCACGTGGGCACACAGTTCCTGAGAG 958
Db 546 GCGCGAGATGGCGGCTTGTACTGGCGCTGGTATTCACGTGGGCACAGTTCCTGAGAG 605
Qy 959 GGAGAGGATACGGATGCAATCCCCCGTTCGATGTCTCAGGTCTCCAAATGGCTGGTG 1018
Db 606 GGACGCGAGTGGCGGCTTGTACTGGCGCTGGTATTCAGGTCTCCAACTATGGCTGGGG 665
Qy 1019 GCTGTTGCTGTGATGAAGAAATTTTGAATATCTGGGTCTTATGATACAGAAATGGA 1078
Db 666 CTTGTTGCTGTGATGAAGATATTTTGAATATCTGGGTCTTACGATACAGAAATGGA 725
Qy 1079 AGTTTGGGAGGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGGGTCT 1138
Db 726 AGTCTGGGAGGAAACCTTGAGTCTCTTTAGGATCTGGCAGTGTGGGCACTCT 785
Qy 1139 GGAACACACCCATGTTCCCATGTGTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCG 1198
Db 786 GGAACGACCCGCTGCTCCCATGTGGCCAGCTCTTCCCTAAGCAAGCTCCCTATTCGCG 845
Qy 1199 CAACAGGCTCTGGCCAAACAGTGTGTGCGAGCTGAAGTATGATGATGAATTTAAGA 1258
Db 846 TAGCAAGGCCCTGGCCAAACAGTGTCCGAGCTGCAGAAAGTGTGATGAATTTAAGA 905
Qy 1259 GCTTCTACTACATCGCAACCCCGTGGCGCTTCCGAACCTTTTGGGATGTGACAGAGAG 1318
Db 906 ACTTCTACTACACCGCAATCCCGAGCCGCTCGAACCTTTGGGAGCTGACAGAGAG 965
Qy 1319 GAAGCAGCTCCGGGCAAGCTCCAGTGTAAAGACTTCAAGTGGTCTTTGGAGACTGTGA 1378
Db 966 GAAGAGCTTCCGGCTAAGCTCCAGTGTAAAGACTTCAAGTGGTCTTTAGATACAGTGA 1025
Qy 1379 TCCAGAACTGATGTGCTGAGGACAGGCTGGCTTCTTCGGGATGTCTCCAGAAACAAGG 1438
Db 1026 CCCAGAAATGCAAGTGGCCAGAGGACAGGCTGGCTTCTTTGGGATGTCTCCAGAAACAGAG 1085
Qy 1439 ACTAACAGACTACTGCTTTGACTATAACCTCCCGATGAACACAGATTTGGGACACCA 1498
Db 1086 CCTAAGAGGACTGCTCTTGACTACATCTCCCAATGAACCACTGTTGAAGGCCACCA 1145
Qy 1499 GGTCACTCTGACTCTGCTGATGGATGGGCGAGAACTCAGTGTTCGAGTACAGTCCCA 1558
Db 1146 GGTCTTCTGTACCTCTGCTGATGGATGGGTCAGAACCACTTTTCGAGTATACGACCG 1205
Qy 1559 GAAAGAAATACGCTATTAACACCCACCGCTGAGGCTGCATTTGCTGTGGAGCAGGAAT 1618
Db 1206 GAAAGAAATACGCTATTAACACCCACCGCTGAGGCTGCATCCAGTGGAGGACGGAA 1265
Qy 1619 GGATACCTTATCATGCTCTCTGCGAAGAACTGCCCCAGAGATCAGAGTTCATCTT 1678
Db 1266 GGATACCTTGTCTGAGTCTCTGCGAGAGACCGTCCAGAGAACACAGGAGTTCATCT 1325
Qy 1679 GCAGGAGGATGGATCTTTTATTTTACGAACAGTCCAAAGAAATGTGTCCAGGCTGGCGAA 1738
Db 1326 ACAGAGGATGGCACTTTAGTTTCAACAGCAGCAGGAAATGTGTGGAGGCCACAGAGAA 1385

Qy 1739 GGAGTCCAGTACAGTTCCTTCCACTTTACGAGACTGCACCAACTCGGATCATCAGAA 1798
Db 1386 GGTGTTAGACACGGCTTTGACCACTACTTACGGACTCTGACCACTCAGATAACCCAGAG 1445
Qy 1799 ATGGTTCCTTCAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGGAGCCCATCGAAGAGGAG 1858
Db 1446 GTGGTTCCTTCAAGGAGCGCATGTCGTAGTGCATCGCT-----TGCCAAGGAAGGAG 1497
Qy 1859 ACTGTGGAGCGGAGGACTCTGCCCAACAAAGACTTAGCTAAGCAGTGACCAACCCACCA 1918
Db 1498 CCTCCACGGCTGTGGACTGTGTGCCAA----- 1526
Qy 1919 AAAAATAGCTGCATGCTTTTGAAGAGGCAATCATTTGCACTTTGTTGCAAGTGTGTGTTG 1978
Db 1527 -----TGAGATGGACAGCGTACACGGTGTCTTCGTGAA---TGTCGTG 1568
Qy 1979 GATTTAGTAAAAATGTAATAAGCTTTGTACTTATTTAGAGAACTTTTAAATGTTCCAA 2038
Db 1569 GGTGTCAAGAAGTGTGAATAGCTTTGTACAGATTTTGAAGAACTTTAAAAATGTCTCCT 1628
Qy 2039 AATACCTATTTTCAAGGGTAACTGTAAGATGTTAACCTTGGTATTTAGAAATTTAA 2098
Db 1629 GGCATCTCTGCTTCTCAGGGCAATCTTAAGATGTT-GCCTTTGGTATTTGGA----- 1680
Qy 2099 ACCTTATAATATTTTCTATCAARAWAWATTTTACAGTCTGTCCTTTTACTCTCATTA 2158
Db 1681 -----AGGATATTTTCTTACAGATGTCATTTACATTTGTCCTTTTATTTCTCATTA 1735
Qy 2159 GCAAAAAAGATAAGATTTTATTTTGGTATTTTACAAGAA 2197
Db 1736 GCAAAAAAGTAAATACTTTTATTTTGATACAAAAA 1774

RESULT 5

BC024475 Mus musculus, clone IMAGE:4976620, mRNA. linear HTC 04-MAR-2003

LOCUS BC024475

DEFINITION BC024475.1 GI:22137472

ACCESSION BC024475.1

VERSION BC024475.1

KEYWORDS HTC.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 2064)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (01-MAR-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

REMARK NTH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: cgapbs-1@mail.nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mcdpaxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Series: IRAP Plate: 59 Row: j Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis
This clone has the following problem: retained intron.
Location/Qualifiers

Qy 1979 GATTAGTAAAGATGTAAGCTTTGTACTTATTTTGAGAACTTTTAAATGTTCCAA 2038
 Db 1679 GGTGTCACAGAGTGTGAATAAGCTTTGTACAGATTTTGAAACCTTTTAAATGTTCTCT 1738
 Qy 2039 AATACCCCTATTTTCAAGGGTAATCGTAAGATGTTAACCCCTGTGTTATTGAAATATAA 2098
 Db 1739 GGCATCTCTGCTTTCTGAGGCAATCTTAAGATGTT-GCCTTTGTTATTGGA- 1790
 Qy 2099 ACCTTATAATATTTTCTATCAARAWAWATTTTACAGTCGTGCTCTTACTCTCATTA 2158
 Db 1791 -----AGATATTTTCTTATCAAGATGTCATTTTACATGTCCTTTTATTTCTCATTA 1845
 Qy 2159 GCAAAAAGATAAAGATTTTATTTTGTGTTATTTTCAAGAAATTTCCAGGTACGAAGATATCT 2218
 Db 1846 GCNAAGAGGTAATACTTTTATTTTGTATACATACACAGCTTCTCTGGTGTT----- 1897
 Qy 2219 GCATGGGTGGAATCAGTTTCAAGCAACGTAATCTTGTGATTAATACCTTCAGCTG 2278
 Db 1898 -----GTAGCACATTTTACACAACTGTTAATACCTTCAATG 1934
 Qy 2279 CGGGGTTAAAGTTTCCAGTATAGAGACTGTCACTAGGAACATTTGTTATGATTT--- 2335
 Db 1935 TTGAGTTAGATCTTCTCAGGGGATGAGTGGACACTAGAAATACTGTTGATTTCTTCA 1994
 Qy 2336 --ATTTCAGGTCAATCAGATCTTCTAGATGTTATTTTAAAGAAATGCTTTT 2383
 Db 1995 GGTGAGTGGGTGAGATGCTCTAGATTCATACATTAAGAAATGCTTTT 2044

AK043977 1752 bp mRNA linear HTC 05-DEC-2002
 Mus musculus 10 days neonate cortex cDNA, RIKEN full-length
 enriched library, clone:A830062E06 product:hypothetical Glycosyl
 transferase, family 2 containing protein, full insert sequence.

AK043977
 VERSION AK043977.1 GI:26090099
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 2 Carninci, P. and Hayashizaki, Y.
 3 Itoh, M., Konno, H., Okazaki, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)

20499374
 MEDLINE 11042159
 PUBMED 10349636

REFERENCE
 1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 2 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)

20499374
 MEDLINE 11042159
 PUBMED 10349636

REFERENCE
 1 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 2 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multipillar sequencer
 Genome Res. 10 (11), 1757-1771 (2000)

20530913
 MEDLINE 11076861
 PUBMED 11076861

REFERENCE
 1 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamahaka, I.,

Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
 Kadota, K., Matsuo, H., Ashburner, M., Batalov, S., Casavant, T.,
 Fleidmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
 Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
 Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,
 Wagner, L., Washio, T., Sakai, K., Okido, F., Furuno, M., Aono, H.,
 Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
 Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
 Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
 Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
 Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
 Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
 Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,
 Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.
 and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)
 21085660
 MEDLINE 11217851
 PUBMED 11217851

REFERENCE
 5 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

6 (bases 1 to 1752)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, K., Saiton, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akai, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.

Direct Submission
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

COMMENT
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/

FEATURES
 Location/Qualifiers
 1..1752
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:A830062E06"
 /db_xref="taxon:10090"
 /clone="A830062E06"
 /tissue_type="cortex"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="10 days neonate"
 757..1612
 /note="hypothetical Glycosyl transferase, family 2
 containing protein (InterPro|IPR001173, evidence:
 InterPro)
 putative"

misc_feature
 371 a 468 c 566 g 347 t

BASE COUNT
 ORIGIN

Query Match 33.1%; Score 943.8; DB 11; Length 1752;
Best Local Similarity 74.4%; Pred. No. 8.8e-94;
Matches 1333; Conservative 0; Mismatches 272; Indels 186; Gaps 4;

QY 54 AGATCGCTGGCTGAGTGGCGGCGCATGTGGGGGCGCACGGCGCGGCGCGCTGCCG 113
DB 24 AGCGCGGGGTGAGTACGGGTGATGTGGGGGCGTGGGTGAGAGGCGCTGCCG 83

QY 114 CGGAACTGCGGCGCGCGCGGAGGCGCTGTGTGCTCTCTGGCGCTACTGTGGGTTGGCC 173
DB 84 CGGGGCTGCGGCGCGCGCGGAGGCGTTG-----CTGGCACTGCTGGCCCTGGCA 134

QY 174 GGGCTGGGCTGCTGCTGCGGCGCGAGCTGTGGGCGCGGCGCGGCGCTGCCGAGCGGA 233
DB 135 GGGCTGGGCGGCTGCTGCGGCG-----TCGAAGCAGGTGCGGGGACGCTGCGACCGGGA 188

QY 234 CCGCGCGGCGCGCGCGCGCGCGCGGCGCGGCGCTCATCGCGGCGCGCGGCTGCCG 293
DB 189 CCGCGCGGCGCTCTCTCCCGCGGCGTCAAGCGAGCTGTGCCAGGCGCGCGTGCC 248

QY 294 CGAAACGCTGGCGCGCGCGGCGCGGCGGCTGCGGCTGCGAGCTGCGAGGCGGAGAGCTG 353
DB 249 GCGATGCTCTGGGAGCGCATGGGAGCGGTGCGGCTGCGAGCTGCGAGGCGGAGAGCTG 308

QY 354 GGGCTGAGGAGAGAGCGGTGCGGCTGCAACAGATTAACTATCTACCTACGCGACCGGATC 413
DB 309 GGGCTGAGGAGAGAGCGGTGAAGCAGCACAGATCAACATCTATCTGAGCGACCGGTATC 368

QY 414 TCACTGACCGCGCGCTGCGGCGCGGCGGCGGCGGCGGCTGCGAAAGAGAGAAATATGAT 473
DB 369 TCGCTGACCGCGGCGCTGCGGCGGCGGCTGGAACCGCT----- 406

QY 474 TATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATAAGAGCTGTGTCAACT 533
DB 407 ----- 406

QY 534 CTCTTGGGACAGTTTACAGTGTCTTGGAGCATCCCCGATATCTGCTAGAAGAGTG 593
DB 407 ----- 406

QY 594 ATCTTTGTAGTACTAGATAGAGACCTTGAGAGCGCTTGGCAATGAGCTT 653
DB 407 -----AGCACCTAAGAAACGCTTGGCCAAAGAGCTG 438

QY 654 TCGGAGCTGCCAGGTGCGCTGATCCGCGCCAAAGAGAGAGGCGCTGGTGGAGCC 713
DB 439 TCACAGCTCCCAAGGTGCGCTGATCCGTGCTAGCAGGAGAGGCGCTAGTGGAGCC 498

QY 714 CGGCTGCTGGGCGGCTGCGGCGAGGCGGATGTTCTGACCTTCTTGGACTGTCACTGT 773
DB 499 CGGCTACTGGGAGCTCTGCGGCGAGGCGAAGTGTGCTGACCTTCTTAGACTGTCACTGT 558

QY 774 GAGTCCAGAGGCTGGCTGGAGCGCTCTGCGAGGATCCATGAAGAGAGTGGCA 833
DB 559 GAATGTGATAGAGGGTGGCTGGAGCGCTCTGCTGAGAGGATCCACGAGAGAGTCCGCA 618

QY 834 GTGGTGTGCCGCGGTGATGATGATCGACTGGAACACCTTCGAATACCTGGGGAAGTCC 893
DB 619 GTGCTGTGCCCTGTATCGATGATGATGATGATGATGATGATGATGATGATGATGATG 678

QY 894 GGGAGCGCCAGATCGCGGCTTTCGACTGGAGGCTGGTGTTCACGTGGCGCACAGATTCT 953
DB 679 GGAGAGCGGAGATGGCGGCTTTCGACTGGCGGCTGGTGTTCACGTGGCGCGGTTCC 738

QY 954 GAGAGGAGAGGATAGGATGCAATCCCGCGTGTGATGATGATGATGATGATGATGATG 1013
DB 739 CAGCGGAGCGGAGTGCATGCGGCTGCGGCGGCTGATGATGATGATGATGATGATGATG 798

QY 1014 GGTGGGCTGTTTCTGTGAGTAAGAATAATTTTGAATATCTGGGCTCTTATGATACAGGA 1073
DB 799 GGGGGCTGTTTCTGTGAGTAAGAGATATTTTGAATATCTGGGCTCTTACGATACAGGA 858

QY 1074 ATGGAATTTTGGGAGAGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGGG 1133
DB 859 ATGGAATCTGGGAGAGAGAAACCTTGTAGTTCTCTTTAGGATCTGGCAGTGTGGGG 918

QY 1134 GTTCTGGAACACACCATGTTCCCATGTTGGCCATG-TTTTCCCAAGCAAGCTCCCTA 1192
DB 919 ACTTCTGGAACCGCACCGTCTCCCATGTTGGGCCACGCTTCTTCCCTAAGCAAGCTCCCTA 978

QY 1193 CTCCCGCAACAAGGCTCTGGCCAAACAGTGTTCGTGAGCTGGAAGTATGATGATGAAT 1252
DB 979 TTCCCGTAGCAAGGCTCTGGCCAAACAGTGTTCGAGCTGCAAGTGTGGATGATGAAT 1038

QY 1253 TAAAGAGCTCTACTACATCGCAACCCCGTGGCCGCTTGGAAACCTTTTGGGAGTGTGAC 1312
DB 1039 TAAAGAACTCTACTACACCGCAATCCCGAGGCGCGCTGGAAACCTTTTGGGAGTGTGAC 1098

QY 1313 AGAGAGGAGAGCTCCGGGAGCAAGCTCCAGTGTGAAGCTTCAAGTGTTCCTTGGAGAC 1372
DB 1099 AGAGAGGAGAGCTTCCGGCTAAGCTCCAGTGTGAAGCTTCAAGTGTTCCTTGGAGAC 1158

QY 1373 TGTGTATCCAGAACTGCTATGCTGAGGACAGGCTGGCTTCTTGGGATGCTCCAGAA 1432
DB 1159 AGTGTACCCAGAACTGCTGAGGACAGGCTGGCTTCTTGGGATGCTCCAGAA 1218

QY 1433 CAAAGAGCTAAACAGACTATGCTTTGACTATTAACCTCCCGATGAAACAGATTTGGG 1492
DB 1219 CAGAGGCTAAGAGGCTGCTGCTTGAATCAATCTCCCAATGAAACCCATGTTGAAG 1278

QY 1493 ACACAGGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1552
DB 1279 CCACAGGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1338

QY 1553 GTCCAGAGAGAAATACGCTATAACACCAACGAGCTGAGGCTGCTGCTGCTGCTGCTGCT 1612
DB 1339 GACCCGAGAGAAATACGCTATAACACCAACGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1398

QY 1613 AGAATGAGTACCTTTATCATGCTCTGCGAAGAACTGCGCCAGAGAACTCAGAACTT 1672
DB 1399 CGGAGGAGTACCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1458

QY 1673 CATCTTGGAGGAGTGGATCTTTATTTTCAAGAGCTGAGGCTGCTGCTGCTGCTGCTGCT 1732
DB 1459 CATCTTACAGGAGTGGATGCTTTAGTTTCAAGCAAGAGGAGTGTGCTGCTGCTGCTGCT 1518

QY 1733 GAGGAGGAGTGGAGTGGATGCTTTGCTTCCACTTTCAGAGCTGCTGCTGCTGCTGCTGCT 1792
DB 1519 AGAAGAGTGTGTAGCAACGGCTTTGCAACATCTTACGGAGTGTACCACTCAGATA 1578

QY 1793 TCAGAAATGTTTCTTCAAGAGCGCATGTTATGAAGCTTCTGCTGCTGCTGCTGCTGCT 1843
DB 1579 CCAGAGGTTGTTCTTCAAGAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1629

RESULT 7
CD251293
LOCUS
DEFINITION
IMAGE:30384460 5', mRNA sequence.
CD251293
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CD251293 909 bp mRNA linear EST 22-MAY-2003
AGENCOURT 14211744 NIH.MGC.179 Homo sapiens cDNA clone
IMAGE:30384460 5', mRNA sequence.
CD251293.1 GI:31011759
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 909).
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein
 cDNA Library Preparation: Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: NDAM449 row: p column: 05
 High quality sequence stop: 701.
 Location/Qualifiers
 1..909
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30384460"
 /tissue_type="Pituitary"
 /lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
 /clone_lib="NIH_MGC_179"
 /notes="Organ: Brain; Vector: pCMV-SPORT6.1; Site 1: EcoRV
 (destroyed); Site 2: NotI; Library is oligo-dr primed and
 directionally cloned (EcoRV site is destroyed upon cloning
). Average insert size 1.1 kb. Library was constructed by
 (Invitrogen). Note: this is a NIH_MGC Library."

FEATURES

source

BASE COUNT 273 a 178 c 198 g 260 t

ORIGIN

Query Match 30.5%; Score 868; DB 14; Length 909;
 Best Local Similarity 98.3%; Pred. No. 2e-85;
 Matches 871; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

1490 GGGACACAGGTCTTCTGACCTCTGTCATGGATGGCGGAGATCAGTTTCGAGTA 1549
 |||||
 22 GGGATCCAGGTCTTCTGACCTCTGTCATGGATGGCGGAGATCAGTTTCGAGTA 81

1550 CAGCTCCAGAGAAATAGCTATACACCCAGCTGAGGCTGTCATGCTGTGGA 1609
 |||||
 82 CAGCTCCAGAGAAATAGCTATACACCCAGCTGAGGCTGTCATGCTGTGGA 141

1610 AGCAGGAATGGATACCTTATCATGCTCTGCGAAGAAATGCCCCAGAGATCAGAA 1669
 |||||
 142 AGCAGGAATGGATACCTTATCATGCTCTGCGAAGAAATGCCCCAGAGATCAGAA 201

1670 GTTCATCTTGAGAGATGATCTTTATTTTCAGAACAGCTCCAGAAATGTCTCAGGC 1729
 |||||
 202 GTTCATCTTGAGAGATGATCTTTATTTTCAGAACAGCTCCAGAAATGTCTCAGGC 261

1730 TGGCAGGAGAGTGGAGTACAGTTGTTTCCACTCTTACGAGACTGCACCACTCGGA 1789
 |||||
 262 TGGCAGGAGAGTGGAGTACAGTTGTTTCCACTCTTACGAGACTGCACCACTCGGA 321

1790 TCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGAAGCTCTGTATCAAGAGCCCA 1849
 |||||
 322 TCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGAAGCTCTGTATCAAGAGCCCA 381

1850 TCGAAGGAGACTGTGGAGCCAGGACTGTGCCCAACAAAGACTTAGTAAAGAGTACGAC 1909
 |||||
 382 TCGAAGGAGACTGTGGAGCCAGGACTGTGCCCAACAAAGACTTAGTAAAGAGTACGAC 441

1910 AACCACCAAAACTAGGCTGCTGTTTTCGAAGGCAATCATTTTGCCATTTGTGAAA 1969
 |||||
 442 AACCACCAAAACTAGGCTGCTGTTTTCGAAGGCAATCATTTTGCCATTTGTGAAA 501

1970 GTTGTGTTGGATTTAGTAAAAATGTGAAATAAGCTTTGTACTTATTTTGAGAACTTTTAA 2029
 |||||
 502 GTTGTGTTGGATTTAGTAAAAATGTGAAATAAGCTTTGTACTTATTTTGAGAACTTTTAA 561

2030 ATGTTCCAAATACCTATTTTCAAAGGTAATCGTAAGATGTTAACCTTTGGTATTAG 2089
 |||||
 562 ATGTTCCAAATACCTATTTTCAAAGGTAATCGTAAGATGTTAACCTTTGGTATTAG 621

2090 AAAATTAACCTTATATATTTTCTATCAARAWAWATTTTACAGTCGTCCTTTTA 2149
 |||||

Db 622 AAAATTAACCTTATATATTTTCTATCAAGATGTATATTTTACAGTCGTCCTTTTA 681

Qy 2150 CTCTCATTAGCAAAAAGATAAGATTTTATTTTGTATTTTCAAGAAATCCAGGTACG 2209
 |||||

Db 682 CTCTCATTAGCAAAAAGATAAGATTTTATTTTGTATTTTCAAGAAATCCAGGTACG 741
 |||||

Qy 2210 AAGATATCTGCATGGGTGGAATCAGGTTCAAGCAACGTACTTTGCATTAACTGATAA 2269
 |||||

Db 742 AAGATATCTGCATGGGTGGAATCAGGTTCAAGCAACGTACTTTGCATTAACTGATAA 801
 |||||

Qy 2270 CTTAGCTCGGGGTTAAAGTTTCCAGTATAGAGAGACTGTCTACTAGAACATTTGAT 2329
 |||||

Db 802 CTTAGCTCGGGGTTAAAGTTTCCAGTATAGAGAGACTGTCTACTAGAACATTTGAT 861
 |||||

Qy 2330 TGATTTTTCAGGTCTATTCAGATCTTCTAGATGTTATTTTAAAGA 2375
 |||||

Db 862 TGATTTTTCAGGTCTATTCAGATCTTCTAGATGTTATTTTAAAGA 907
 |||||

RESULT 8

CB243715/c

LOCUS

DEFINITION

CB243715

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

CB243715 745 bp mRNA linear EST 12-FEB-2003
 UI-CF-FNO-9gf-d-02-0-UI.s1 UI-CF-FNO Homo sapiens cDNA clone
 UI-CF-FNO-9gf-d-02-0-UI 3', mRNA sequence.

CB243715
 EST.
 Homo sapiens (human)

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 745)
 Ronaldo,M.F., Lennon,G. and Soares,M.B
 Normalization and subtraction: two approaches to facilitate gene
 discovery

Genome Res. 6 (9), 791-806 (1996)
 97044477
 8889548

Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171

Email: paul-mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com) or from Open Biosystems
 (www.openbiosystems.com).

Seq primer: M13 FORWARD
 POLYA=Yes.

Location/Qualifiers
 1..745
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-FNO-9gf-d-02-0-UI"
 /tissue_type="Human Lung Epithelial cells"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-CF-FNO"
 /notes="Organ: Lung; Vector: p7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-CF-FNO is a subtracted cDNA library derived from two
 normalised Human lung epithelial cell libraries (EN1 and
 DU1) The library was subtracted according to according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. For additional information, contact:
 bento-soares@uiowa.edu

TAG LIB=UI-CF-FNO
 TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS

```
6hr to LPS 24h
TAG SEQ=CTGCTCAGGT"
BASE COUNT 249 a 132 c 115 g 249 t
ORIGIN
Query Match 25.9%; Score 738.2; DB 14; Length 745;
Best Local Similarity 98.9%; Pred. No. 2.8e-71;
Matches 737; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
Qy 2088 AGAAATTTAAACCTTATAATATTTTCTATCAARAWAWATTTTACAGTCGTCGCTTT 2147
Db 745 AGAAATTTAAACCTTATAATATTTTCTATCAAGATGTATTTTACAGTCGTCGCTTT 686
Qy 2148 TACTCTCATAGCAAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTCACAGTA 2207
Db 685 TACTCTCATAGCAAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTCACAGTA 626
Qy 2208 CGAAGATATCTGATGGTGGAAATCAGGTTCAAGCAACGCTCTTCCATTAACGTATAA 2267
Db 625 CGAAGATATCTGATGGTGGAAATCAGGTTCAAGCAACGCTCTTCCATTAACGTATAA 566
Qy 2268 TACCTCAGTCGCGGGTTAAAGTTTCCAGTATAGAGAGACTCTCACTAGGAACATTTGT 2327
Db 565 TACCTCAGTCGCGGGTTAAAGTTTCCAGTATAGAGAGACTCTCACTAGGAACATTTGT 506
Qy 2328 ATTGATTTATTCAGTCAATGAGATCTTCTAGATGTATTTTAAAGAAATGCTTTTGGT 2387
Db 505 ATTGATTTATTCAGTCAATGAGATCTTCTAGATGTATTTTAAAGAAATGCTTTTGGT 446
Qy 2388 TATGTGTGTACACAGTTTAAACACTCATAATGTTTCTATGTCAGCCAAAGAGACTAAC 2447
Db 445 TATGTGTGTACACAGTTTAAACACTCATAATGTTTCTATGTCAGCCAAAGAGACTAAC 386
Qy 2448 AAAGCTGAATCTCAGAGAAACAAATTTGCTTACTAAGCTGAGTCAACTTGAGACGAACT 2507
Db 385 AAAGCTGAATCTCAGAGAAACAAATTTGCTTACTAAGCTGAGTCAACTTGAGACGAACT 326
Qy 2508 TCTAACAATCGCGCACTGTAGTGTGGTGTCTTACCACATATGACTTTTAAACATGTTTA 2567
Db 325 TCTAACAATCGCGCACTGTAGTGTGGTGTCTTACCACATATGACTTTTAAACATGTTTA 266
Qy 2568 TATCATTTTAAATTTTATGATACGGTAGTGTGAGGAGAAATGTAATGTTCTATATGAA 2627
Db 265 TATCATTTTAAATTTTATGATACGGTAGTGTGAGGAGAAATGTAATGTTCTATATGAA 206
Qy 2628 ATTCTTTTCAAGTTTGTTCATTAATAACAGTTTATTAATTTAAATCAGGCTTAGAGTTT 2687
Db 205 ATTCTTTTCAAGTTTGTTCATTAATAACAGTTTATTAATTTAAATCAGGCTTAGAGTTT 146
Qy 2688 GTGCTGCTCAACTGCTGTGAAAATTTCTCTGAGTAAATTTCTGATTTGTAATGATCCAG 2747
Db 145 GTGCTGCTCAACTGCTGTGAAAATTTCTCTGAGTAAATTTCTGATTTGTAATGATCCAG 86
Qy 2748 ACCAACCCTGAGATTTTGTCAACCTGATTAAGTCAATATGAATGATTAAGATGTGAG 2807
Db 85 ACCAACCCTGAGATTTTGTCAACCTGATTAAGTCAATATGAATGATTAAGATGTGAG 26
Qy 2808 AACAAAAAATAAAAAAAAAAAAAA 2832
Db 25 AACACTGAAAAAATAAAAAAAAAAAAAA 1
RESULT 9
BM978800/c
LOCUS
DEFINITION UI-CF-DUI-1-adp-g-07-0-UI.61 UI-CF-DUI Homo sapiens cDNA clone
BM978800
VERSION BM978800.1 GI:19598602
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 745)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com)
Seq primer: M13 FORWARD
POLYA=Yes.

Location/Qualifiers
1. 745
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-DUI-1-adp-g-07-0-UI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-DUI"
/note="Organ: Lung; Vector: p773-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-DUI is a normalized cDNA library containing the
following tissue(s): Primary Lung Epithelial Cells The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into p773-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GGCTGTAGGC.
TAG LIB=UI-CF-DUI
TAG TISSUE=Lung Epithelial Cells Tissue nos 359-368
TAG SEQ=GGCTGTAGGC"

BASE COUNT 248 a 132 c 117 g 248 t
ORIGIN

Query Match 25.4%; Score 723; DB 12; Length 745;
Best Local Similarity 98.5%; Pred. No. 1.2e-69;
Matches 734; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Qy 2089 GAAAAATTAACCTTATAATATTTTCTAT-CAARAWAWATTTTACAGTCGTCGCTTT 2147
Db 745 GAAAAATTAACCTTATAATATTTTCTATCAAGATGTATTTTACAGTCGTCGCTTT 686
Qy 2148 TACTCTCATAGCAAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTCACAGTA 2207
Db 685 TACTCTCATAGCAAAAGATAAGATTTTATTTTGGTATTTTACAGAAATTCACAGTA 626
Qy 2208 CGAAGATATCTGATGGTGGAAATCAGGTTCAAGCAACGCTCTTCCATTAACGTATAA 2267
Db 625 CGAAGATATCTGATGGTGGAAATCAGGTTCAAGCAACGCTCTTCCATTAACGTATAA 566
Qy 2268 TACCTCAGTCGCGGGTTAAAGTTTCCAGTATAGAGAGACTGTCTAGGAACATTTGT 2327
Db 565 TACCTCAGTCGCGGGTTAAAGTTTCCAGTATAGAGAGACTGTCTAGGAACATTTGT 506

QY 2328 ATTGATTTATTCAGTCAATGAGATCTCTAGATGTTATTTTAAAAAGAAAGTCTTTTGGT 2387
Db 505 ATTGATTTATTCAGTCAATGAGATCTCTAGATGTTATTTTAAAAAGAAAGTCTTTTGGT 446
QY 2388 TATGTTGTTGTTACACAGTAACTCACTCAATGTTTCATGTCAGCCAAAGAGGACTAAC 2447
Db 445 TATGTTGTTGTTACACAGTAACTCACTCAATGTTTCATGTCAGCCAAAGAGGACTAAC 386
QY 2448 AAAGCTGAATCTCAGAGAACAAATTTGCTTTTAACTGAGTCAACTTGAAGCGAACT 2507
Db 385 AAAGCTGAATCTCAGAGAACAAATTTGCTTTTAACTGAGTCAACTTGAAGCGAACT 326
QY 2508 TCTAACAAATCGGCACCTGAGTGGCTGCTTACCACTATGACTTTTAAAAAGATGTTTA 2567
Db 325 TCTAACAAATCGGCACCTGAGTGGCTGCTTACCACTATGACTTTTAAAAAGATGTTTA 266
QY 2568 TATCAATTTTAAATTTTATGATAGCTGAGTGTCTAGGAGAAATGTAATGTTCTATATGAA 2627
Db 265 TATCAATTTTAAATTTTATGATAGCTGAGTGTCTAGGAGAAATGTAATGTTCTATATGAA 206
QY 2628 ATTCTTTTCAAGTTTGTTCATTAATAACAGTTTATTAATTAATCAGCGTTAGAGTTT 2687
Db 205 ATTCTTTTCAAGTTTGTTCATTAATAACAGTTTATTAATTAATCAGCGTTAGAGTTT 146
QY 2688 GTGCTGCTGCACTGCTGTGAAATTTCTCTGAGTAAATCTGATTTCTGATGATCCAG 2747
Db 145 GTGCTGCTGCACTGCTGTGAAATTTCTCTGAGTAAATCTGATTTCTGATGATCCAG 86
QY 2748 ACCAACCTGAGATTTTGTCAACCTGATTAAGTCAATATGAAATGATTAATAAGAGATGTGAG 2807
Db 85 ACCAACCTGAGATTTTGTCAACCTGATTAAGTCAATATGAAATGATTAATAAGAGATGTGAG 26
QY 2808 AACAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAA 2832
Db 25 AACACTGAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAA 1

RESULT 10
BU854444
LOCUS
DEFINITION AGENCOURT_10409743 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6621725
5' mRNA sequence.
ACCESSION BU854444
VERSION BU854444.1 GI:24039410
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 782)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2873 row: o column: 05
High quality sequence stop: 634.
Location/Qualifiers
1..782
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6621725"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_82"

FEATURES
source

/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgctggcc); Site_2: SfiI (ggcgctggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCATATGACC-3' and 3' adaptor sequence:
5'-ATCTAGAGCCGAGCGCGGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."

BASE COUNT	244 a	123 c	142 g	271 t	2 others
ORIGIN					
Query Match	25.2%	Score 718.6;	DB 13;	Length 782;	
Best Local Similarity	98.5%;	Pred. No. 3.6e-69;			
Matches 729;	Conservative	5;	Mismatches	5;	Indels 1;
Gaps	1;				
QY 1899	GCAGTGCACGAAACCCACCAAAACCTAGGCTGCTGCTTTGTAAGAGGCAATCATTTTGC	1958			
Db 3	GCAGTGCACGAAACCCACCAAAACCTAGGCTGCTGCTTTGTAAGAGGCAATCATTTTGC	62			
QY 1959	CATTTGTGAAGTTGTTGGATTAGTAAATGTAATGTAATGTAATGTAATGTAATGTAATGTA	2018			
Db 63	CATTTGTGAAGTTGTTGGATTAGTAAATGTAATGTAATGTAATGTAATGTAATGTAATGTA	122			
QY 2019	GAACTTTTAAATGTTCCAAATACCTATTTTCAAAGGGTAATCGTAAGATGTTAACCC	2078			
Db 123	GAACTTTTAAATGTTCCAAATACCTATTTTCAAAGGGTAATCGTAAGATGTTAACCC	182			
QY 2079	TTGGTATTTAGAAAAATTAACCTTATTAATTTTCTATCAARAWAATTTTACAGT	2138			
Db 183	TTGGTATTTAGAAAAATTAACCTTATTAATTTTCTATCAARAWAATTTTACAGT	242			
QY 2139	CGTGCCCTTTTACTCTCATTTAGCAAAAAGATAAGATTTTATTTTGGTATTTTCAAGAAT	2198			
Db 243	CGTGCCCTTTTACTCTCATTTAGCAAAAAGATAAGATTTTATTTTGGTATTTTCAAGAAT	302			
QY 2199	TCCAGGTAGCAAGATATCTGCATGGTGGAATTCAGTTTCAAGCAACGTACTTTGCATT	2258			
Db 303	TCCAGGTAGCAAGATATCTGCATGGTGGAATTCAGTTTCAAGCAACGTACTTTGCATT	362			
QY 2259	AACGTAAATACCTCAGCTCGGGGTTAAAGTTTCCAGTATAGAGAGACTGTCACCTAG	2318			
Db 363	AACGTAAATACCTCAGCTCGGGGTTAAAGTTTCCAGTATAGAGAGACTGTCACCTAG	422			
QY 2319	GAACATTTGATTTGATTTTATTTTCAAGTCAATTCAGATGTTTATTTTAAAGAAATG	2378			
Db 423	GAACATTTGATTTGATTTTATTTTCAAGTCAATTCAGATGTTTATTTTAAAGAAATG	482			
QY 2379	CTTTTGGTTATGTTGTTTACCACTGTAACCTCCATTAATGTTTCATGTCAGCCAAAGA	2438			
Db 483	CTTTTGGTTATGTTGTTTACCACTGTAACCTCCATTAATGTTTCATGTCAGCCAAAGA	542			
QY 2439	GGACTAACCAAGAGCTGAAATCTCAGAGAACAAATTTGCTTTTACTTAAGCTGAGTCACTGA	2498			
Db 543	GGACTAACCAAGAGCTGAAATCTCAGAGAACAAATTTGCTTTTACTTAAGCTGAGTCACTGA	602			
QY 2499	GAGGAACTTCTAACAATGCGGCACTGAGTGGCTGGTCTTACCACTATGACTTTTAA	2558			
Db 603	GAGGAACTTCTAACAATGCGGCACTGAGTGGCTGGTCTTACCACTATGACTTTTAA	662			
QY 2559	ACATGTTTATATCATTTTATTTTATGATACCGTAGTGTGAGGAGAAATGTAAT-GT	2617			
Db 663	ACATGTTTATATCATTTTATTTTATGATACCGTAGTGTGAGGAGAAATGTAATGTT	722			
QY 2618	TCTATATGAATTCCTTTT 2637				
Db 723	TCTATATGAATTCCTTTT 742				

RESULT 11
BM977212/c
LOCUS BM977212

711 bp mRNA linear EST 21-FEB-2003

```

DEFINITION  UI-CF-DUI-ads-1-04-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone
ACCESSION   BM977212
VERSION     BM977212.1 GI:19595389
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 711)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
PUBMED      8889548
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com) or from Open Biosystems
            (www.openbiosystems.com).
            Seq primer: M13 FORWARD
            PolyA=yes.

FEATURES             Location/Qualifiers
     1..711
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="UI-CF-DUI-ads-1-04-0-UI"
         /tissue_type="Primary Lung Epithelial Cells"
         /dev_stage="Adult"
         /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
         /clone_lib="UI-CF-DUI"
         /notes="Organ: Lung; Vector: p773-Pac (Pharmacia) with a
         modified polylinker; Site 1: EcoR I; Site 2: Not I;
         UI-CF-DUI is a normalized cDNA library containing the
         following tissue(s): Primary Lung Epithelial Cells The
         library was constructed according to Bonaldo, Lennon and
         Soares, Genome Research, 6:791-806, 1996. First strand
         cDNA synthesis was primed with an oligo-dT primer
         containing a Not I site. Double stranded cDNA was ligated
         to an EcoR I adaptor, digested with Not I, and cloned
         directionally into p773-Pac vector. The oligonucleotide
         used to prime the synthesis of first-strand cDNA contains
         a library tag sequence that is located between the Not I
         site and the (dr)18 tail. The sequence tag for this
         library is GGCTGTAGGC.
         TAG LIB=UI-CF-DUI
         TAG TISSUE=Lung Epithelial Cells Tissue nos 359-368
         TAG_SEQ=GGCTGTAGGC"
BASE COUNT      236 a 129 c 112 g 232 t 2 others
ORIGIN

Query Match      24.8%; Score 705.4; DB 12; Length 711;
Best Local Similarity 98.9%; Pred. No. 1e-67;
Matches 703; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 2115 CTATCAARWRWATTTTACAGTCGGCGCTTTTACTCTCTATTAGCAAAAAGATAAGA 2174
    |||||:::|
Db 711 CTATCAAGATGTATATTTTACAGTCGGCGCTTTTACTCTCTATTAGCAAAAAGATAAGA 652

Qy 2175 TTTTATTTGGTATTTCACAGAAATCCAGGTACAGATATCTGCATGGTGAATCA 2234
    |||||:::|
Db 651 TTTTATTTGGTATTTCACAGAAATCCAGGTACAGATATCTGCATGGTGAATCA 592

```

```

Qy 2235 GGTTCAAGCAACGACTCTTTTCATTAATGATGATTAATACCTCAGCTGCGGGTTAAAGTTTTC 2294
    |||||:::|
Db 591 GGTTCAAGCAACGACTCTTTTCATTAATGATGATTAATACCTCAGCTGCGGGTTAAAGTTTTC 532

Qy 2295 CCAGTATAGAGAGACTGTCTACTAGGACATGTTATGATTTATTCAGGTCAATGAGATCT 2354
    |||||:::|
Db 531 CCAGTATAGAGAGACTGTCTACTAGGACATGTTATGATTTATTCAGGTCAATGAGATCT 472

Qy 2355 TCTAGATGTTATTTTAAAAAGAAATGCTTTTGGTTTATGTTGCTACACAGTTAACTC 2414
    |||||:::|
Db 471 TCTAGATGTTATTTAAAAAGAAATGCTTTTGGTTTATGTTGCTACACAGTTAACTC 412

Qy 2415 CATAATGTTTCATGTCAGCCAAAGAGGACTAACCAAGCTGAAATCTCAGAGAAACAATTTG 2474
    |||||:::|
Db 411 CATAATGTTTCATGTCAGCCAAAGAGGACTAACCAAGCTGAAATCTCAGAGAAACAATTTG 352

Qy 2475 CTTTACTAAGCTGAGTCAACTTTCAGAGCGAATCTTAACCAATGCCGACCTGTAGTGTGGC 2534
    |||||:::|
Db 351 CTTTACTAAGCTGAGTCAACTTTCAGAGCGAATCTTAACCAATGCCGACCTGTAGTGTGGC 292

Qy 2535 TGGTTCTTACCACATGACTTTTAAAAACATGTTTATATCATTTTAAATTTTATGATACGGT 2594
    |||||:::|
Db 291 TGGTTCTTACCACATGACTTTTAAAAACATGTTTATATCATTTTAAATTTTATGATACGGT 232

Qy 2595 AGTGTCAAGGAGAAATGTAATGTTCTATATGAAATTCCTTTTCAAGTTTGTTCATTAAT 2654
    |||||:::|
Db 231 AGTGTCAAGGAGAAATGTAATGTTCTATATGAAATTCCTTTTCAAGTTTGTTCATTAAT 172

Qy 2655 AACAGTTATTAATTTAAATCAGCGTTAGAGTTTGTGCTGCTGCNACTGCTGTGAAATTT 2714
    |||||:::|
Db 171 AACAGTTATTAATTTAAATCAGCGTTAGAGTTTGTGCTGCTGCNACTGCTGTGAAATTT 112

Qy 2715 CTCTGAGTAATCTCTGATTTCTGAATGATCCAGACCAACCTCAGATTTTGTCAACCTGA 2774
    |||||:::|
Db 111 CTCTGAGTAATCTCTGATTTCTGAATGATCCAGACCAACCTCAGATTTTGTCAACCTGA 52

Qy 2775 TTAAGTCAATGATGAATGATTTAAAAAGATGTGAGAACAAAAAATAAAAAA 2825
    |||||:::|
Db 51 TTAAGTCAATGATGAATGATTTAAAAAGATGTGAGAACAAAAAATAAAAAA 1

```

```

RESULT 12
BM970195/c
LOCUS     BM970195
DEFINITION  UI-CF-EC1-abv-1-07-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone
            UI-CF-EC1-abv-1-07-0-UI 3', mRNA sequence.
ACCESSION  BM970195
VERSION     BM970195.1 GI:19587782
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 705)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
PUBMED      8889548
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research

```

Genetics (www.reegen.com) or from Open Biosystems

(www.openbiosystems.com).

Seq primer: M13 FORWARD

POLYA=Yes.

FEATURES

Location/Qualifiers
1..705
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="UI-CF-EC1-abv-i-07-0-UI"
/tissue_type="Lung"
/lab_host="DHI0B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EC1"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-CF-EC1 is a normalized cDNA library containing the following tissue(s): Normal lung from adult and from fetal day 64, day 87, week 19 and week 42. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (d)18 tail. The sequence tag for this library is AAGTGGCTTAC.

TAG LIB=UI-CF-EC1
TAG TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371 and 380-383
TAG_SEQ=AAGTGGCTTAC"

BASE COUNT 234 a 129 c 111 g 231 t
ORIGIN

Query Match 24.6%; Score 699.8; DB 12; Length 705;
Best Local Similarity 99.4%; Pred. No. 4.2e-67;
Matches 701; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2128 WATTTTACAGTCGTCCTTTTACTCTCATTTAGCAAAAAGATAAAGATTTTATTTGGTA 2187
DB 705 TATTTTACAGTCGTCCTTTTACTCTCATTTAGCAAAAAGATAAAGATTTTATTTGGTA 646

QY 2188 TTTTACAAAGATTTCCAGTAGCAGATATCTGCATGGGTGGAATTCAGGTTCAAGCAACG 2247
DB 645 TTTTACAAAGATTTCCAGTAGCAGATATCTGCATGGGTGGAATTCAGGTTCAAGCAACG 586

QY 2248 TACTTTGCAATTAACGTAAATACCTCAGCTCGGGGTTTAAAGTTTCCAGTATAGAGAG 2307
DB 585 TACTTTGCAATTAACGTAAATACCTCAGCTCGGGGTTTAAAGTTTCCAGTATAGAGAG 526

QY 2308 ACTGTCTACTAGGAACATTTGATTTATTTTACAGTCAATGAGATCTTCTAGATGTATTT 2367
DB 525 ACTGTCTACTAGGAACATTTGATTTATTTTACAGTCAATGAGATCTTCTAGATGTATTT 466

QY 2368 TAAAAGAAATGCTTTTGGTTATGTTGTTGTTACCAAGTTTAACTCATATGTTTCATG 2427
DB 465 TAAAAGAAATGCTTTTGGTTATGTTGTTGTTACCAAGTTTAACTCATATGTTTCATG 406

QY 2428 TCAGCCAAAGAGGACTTAACCAAGCTGAAATCTCAGAGAAATTTGCTTTACTAAGCTG 2487
DB 405 TCAGCCAAAGAGGACTTAACCAAGCTGAAATCTCAGAGAAATTTGCTTTACTAAGCTG 346

QY 2488 AGTCAACTTGAGAGCGAATCTTAACAATGCGCACTGTAGTGGCTGGTTTACCACT 2547
DB 345 AGTCAACTTGAGAGCGAATCTTAACAATGCGCACTGTAGTGGCTGGTTTACCACT 286

QY 2548 ATGACTTTTAAACATGTTTATATCTTTTATGTTTATGATACGGTATGTCAGGGAGA 2607
DB 285 ATGACTTTTAAACATGTTTATATCTTTTATGTTTATGATACGGTATGTCAGGGAGA 226

QY 2608 AATGTAATGTTCTATATGAATTCCTTTTCAAGTTTCTTCAATTAATAACAGTTATTAAT 2667

Db 225 AATGTAATGTTCTATATGAATTCCTTTTCAAGTTTGTTCATTAATAACAGTTATTAAT 166
QY 2668 TTAATTCAGCGTTAGAGTTTGTCTGCTGCAATGCTGTGAAATTTCTCTGAGTAATTC 2727
Db 165 TTAATTCAGCGTTAGAGTTTGTCTGCTGCAATGCTGTGAAATTTCTCTGAGTAATTC 106
QY 2728 TGATTTGTGAATGATCCAGCAACCTCAGATTTTGTCAACCTGATTAAGTCAATATG 2787
Db 105 TGATTTGTGAATGATCCAGCAACCTCAGATTTTGTCAACCTGATTAAGTCAATATG 46
QY 2788 AATGATTTAAAAGATGTGAGAACAAAAAATTTTGTCAACCTGATTAAGTCAATATG 2832
Db 45 AATGATTTAAAAGATGTGAGAACAAAAAATTTTGTCAACCTGATTAAGTCAATATG 1

RESULT 13

AW969574/c

LOCUS

DEFINITION

AW969574

AW969574

AW969574.1

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished

Contact: John Quackenbush

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528

Fax: 301 838 0208

Email: johnq@tigr.org

Plate: 267

Seq primer: Forward.

Location/Qualifiers

1..741

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone_lib="MAGE resequences, MAGK"

/note="Vector: pBluescript-SKm"

BASE COUNT 252 a 135 c 120 g 234 t

ORIGIN

Query Match 23.4%; Score 666.6; DB 10; Length 741;

Best Local Similarity 97.3%; Pred. No. 1.6e-63;

Matches 694; Conservative 4; Mismatches 13; Indels 2; Gaps 2;

QY 2096 AAAACCTTATAATATTTTCTATCAAPAWAWA- TTTTACAGTCGTCCTTTTAC-TCT 2153

Db 713 AAAACCTTATAATATTTTCTATCAAPAWAWA- TTTTACAGTCGTCCTTTTAC-TCT 654

QY 2154 CATTAGCAAAAAGATATAAGATTTTATTTTGGTATTTTACAGAAATTTCCAGGTACGAGA 2213

Db 653 CATTAGCAAAAAGATATAAGATTTTATTTTGGATTTTACAGAAATTTCCAGGTACGAGA 594

QY 2214 TATCTCATGGGTGGAATTCAGGTTTCAAGCAACCTACTTTTGGATTAATCTGATTAATCTC 2273

Db 593 TATCTCATGGGTGGAATTCAGGTTTCAAGCAACCTACTTTTGGATTAATCTGATTAATCTC 534

QY 2274 AGCTGCGGGGTAAAGTTTCCAGATAGAGACTGTCTACTAGGAACATTTGATTTGAT 2333

Db 533 AGCTGCGGGGTAAAGTTTCCAGATAGAGACTGTCTACTAGGAACATTTGATTTGAT 474

QY 2334 TTATTACGGTCAATGAGATCTTCTAGATGTATTTTAAAAAGAAATGCTTTTGGTATGTG 2393

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 19:41:23 ; Search time 21 Seconds
(without alignments)
1170.600 Million cell updates/sec

Title: US-10-074-527-2
Perfect score: 3124
Sequence: 1 MWGRTARRCPRELRRGREA.....LLRDCTNSDHQKWFKERM 581

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1638	52.4	578	4	US-09-217-306B-2
2	1625.5	52.0	560	4	US-09-217-306B-10
3	1142.5	36.6	633	2	US-08-648-298-2
4	1100.5	35.2	559	2	US-08-967-508-9
5	1100.5	35.2	559	3	US-08-967-506-9
6	1100.5	35.2	559.5	5	PCT-US94-02552-9
7	1100	35.2	517	2	US-08-967-508-19
8	1100	35.2	517	3	US-08-967-506-19
9	1100	35.2	517	5	PCT-US94-02552-19
10	1023	32.7	639	3	US-09-376-856-2
11	1017	32.6	638	3	US-09-347-488-2
12	991.5	31.7	603	4	US-09-795-926-43
13	974.5	31.2	506	4	US-09-795-926-31
14	967	31.0	631	4	US-09-795-926-41
15	950	30.4	535	4	US-09-795-926-29
16	697.5	22.3	366	4	US-09-795-926-39
17	680.5	21.8	269	4	US-09-795-926-27
18	450.5	14.4	321	4	US-09-795-926-35
19	433.5	13.9	224	4	US-09-795-926-23
20	256.5	8.2	240	4	US-09-795-926-33
21	234.5	8.1	209	4	US-09-795-926-37
22	239.5	7.7	143	4	US-09-795-926-21
23	237.5	7.6	112	4	US-09-795-926-25
24	155	5.0	1206	4	US-09-252-991A-19632
25	145	4.7	702	4	US-09-437-277-1
26	131.5	4.2	965	4	US-09-437-277-3
27	124.5	4.0	492	2	US-08-468-812-4

28	124.5	4.0	492	2	US-08-468-812-7	Sequence 7, Appli
29	124.5	4.0	492	4	US-08-590-563-4	Sequence 4, Appli
30	124.5	4.0	492	4	US-08-590-563-7	Sequence 7, Appli
31	124.5	4.0	492	4	US-09-770-621-4	Sequence 4, Appli
32	124.5	4.0	492	4	US-09-770-621-7	Sequence 7, Appli
33	112.5	3.6	1618	1	US-07-853-913-4	Sequence 4, Appli
34	112.5	3.6	2860	2	US-08-826-267-2	Sequence 2, Appli
35	109.5	3.5	654	4	US-09-252-991A-22603	Sequence 22603, A
36	108.5	3.5	127	1	US-08-392-828C-39	Sequence 39, Appli
37	108.5	3.5	127	3	US-09-330-945-39	Sequence 39, Appli
38	108.5	3.5	491	2	US-08-468-812-8	Sequence 8, Appli
39	108.5	3.5	491	4	US-08-590-563-8	Sequence 8, Appli
40	108.5	3.5	491	4	US-09-770-621-8	Sequence 8, Appli
41	106	3.4	345	4	US-09-252-991A-26330	Sequence 26330, A
42	105.5	3.4	480	2	US-08-468-812-5	Sequence 5, Appli
43	105.5	3.4	480	4	US-08-590-563-5	Sequence 5, Appli
44	105.5	3.4	480	4	US-09-770-621-5	Sequence 5, Appli
45	104.5	3.3	612	4	US-09-252-991A-30072	Sequence 30072, A

ALIGNMENTS

RESULT 1
US-09-217-306B-2
; Sequence 2, Application US/09217306B
; Patent No. 6465220
; GENERAL INFORMATION:
; APPLICANT: Haasan, Helle
; APPLICANT: Clausen, Henrik
; APPLICANT: Bennett, Eric P.
; TITLE OF INVENTION: Glycosylation Using Galnac-T4 Transferase
; FILE REFERENCE: 8850*1
; CURRENT APPLICATION NUMBER: US/09/217,306B
; CURRENT FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-217-306B-2

Query Match 52.4%; Score 1638; DB 4; Length 578;
Best Local Similarity 56.8%; Pred. No. 1.8e-159;
Matches 324; Conservative 78; Mismatches 150; Indels 18; Gaps 8;
Qy 23 VLLALLALAG-----LGSVLRAORGAGAAEPPTPPRR-----EPVMPRPVPA 72
Db 13 LLLAFLTVAYIFVELLVSTPHASAGAG-ARELGSRRLSDLQKNTEDLSRPLVKKPPADS 71
Qy 73 NALGARGEAVRLQOGEEELRQESVRLHGINIYLSDRISLHRLPERWNPCKEKKVDY 132
Db 72 RALGEGWAKSQUNDELKQOEELRYAINIYLSDRISLHRIEDKRWYCKSQFNY 131
Qy 133 DNLPTSTVIITAFNEAWSTLLRTVYSLTSPDILLEEVILVDYSDREHLKERLANELS 192
Db 132 RLPTSTVIITAFNEAWSTLLRTVYSLTSPDILLEEVILVDYSDREHLKERLANELS 191
Qy 193 GLPKVRLIRANKREGLVRARLLGASARGDVLFTLCHCECHGWLBPPLQRIHEESAV 252
Db 192 NLDRVRLIRANKREGLVRARLLGASARGDVLFTLCHCECHGWLBPPLQRIHEESAV 251
Qy 253 VCVVIDIDWNTTEYLGNSGPOIGGDFWELVTHVTHTVPERIRIMOSPDVIRSPTMAG 312
Db 252 VCVVIDIDWNTTEYLGNSGPOIGGDFWELVTHVTHTVPERIRIMOSPDVIRSPTMAG 311
Qy 313 GLFAVSKKPYEYLGSDYDTGMVWNGENLEFSFRWOCGVLETHPCSHVGHVFPKQAPYS 372
Db 312 GLFAVSKKPYEYLGSDYDTGMVWNGENLEFSFRWOCGVLETHPCSHVGHVFPKQAPYS 371
Qy 373 RNKALANSVRAAEVWMDPEFKELYHHRNPRARLPFGDVTERRKQLRDKLQCKDFKWFLETY 432

Db 372 RPNFLQNTARAAEYVMDEYKEHFYRNPPARKEAYGDISERKLRRLRCKSPDMYLNKV 431
Qy 433 YPELHVPEDEPRGPFGLQNKGLTDYCFDYNPDENQIVGHQVILYLCHGMQNOFFEYTS 492
Db 432 FPNLHVPEDEPRGFWGAIARSISSECLDYNPDNNP-TGANLSLFGCHGQGNOFFEYTS 490
Qy 493 QKEIRYNTHOPEGCIAVEAGMDTLIMHLCEE---TAPENQKFILQEDGSLFHEQSKCVQ 549
Db 491 NKEIRFNS-VTELCAEVPQKQNVGMQNCPRDGPVPANIIWHFKEDGTIFHPHSGGLCLS 549
Qy 550 AARKESSDSFVPLLRDCTNSD-HQKWFPE 578
Db 550 AYRTPEGRPDVQ-MRTCDALDKNQIWSFEK 578

RESULT 2

US-09-217-306B-10
; Sequence 10, Application US/09217306B
; Patent No. 6465220
; GENERAL INFORMATION:
; APPLICANT: Hasean, Helle
; APPLICANT: Bennett, Eric P.
; APPLICANT: Clausen, Henrik
; TITLE OF INVENTION: Glycosylation Using Galnac-T4 Transferase
; FILE REFERENCE: 8850*1
; CURRENT APPLICATION NUMBER: US/09/217,306B
; CURRENT FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: Interferon fragment-Galnac-T4 predicted coding region
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(25)
; OTHER INFORMATION: human interferon signal peptide
; NAME/KEY: MISC FEATURE
; LOCATION: (37)..(560)
; OTHER INFORMATION: Galnac-T4
US-09-217-306B-10

Query Match 52.0%; Score 1625.5; DB 4; Length 560;
Best Local Similarity 59.4%; Pred. No. 3.2e-158;
Matches 309; Conservative 75; Mismatches 129; Indels 7; Gaps 5;

Qy 63 PVMRPPVPANALGARGEAVRLQGEELRLQESVRLHQNIVLSDRISLHRLPERWN 122
Db 44 PLYKKPPADSRALGEGWASKQLNEDLQOEELIERAIVNIVLSDRISLHRIEDKRM 103
Qy 123 PLCKEKYDYNLPRTSVIIAFYNEAWSTLRTVYVLETSFDILLEEVILVDDYSDEH 182
Db 104 YECKSQKFNRTLPSTSVIIAFYNEAWSTLRTVSHVLETSFPAVLKEIILVDDLSRVY 163
Qy 183 LKERLANELSLPKVRLIRANKREGLVRARLLGASARGDVLTFDCHCECHGWLPELL 242
Db 164 LKTOLETYISNLDRLVRLIRTNKREGLVRARLIGATFATGDLVTLFYLCHCECNGLPELL 223
Qy 243 QRIHEESAVCPVDVTDWNTFFVLGNSGEPQIGDFDRLVFTVHTVPERIRMQSPV 302
Db 224 ERIGRYETAVCPVDVTDWNTFFYMQIGDFDRLVFTVHTVPERIRMQSPV 283
Qy 303 DVIRSPMTMAGGLFAVSKKYFYLGSYDTGMVWGGENLEFSPRIWQCQGVLETHPCSHVG 362
Db 284 DPIRSPMTMAGGLFAVSKKYFYLGSYDTGMVWGGENLEFSPRIWQCQGVLETHPCSHVG 343
Qy 363 HVFPKQAPYSRNKALANSVRAAEVWMDFFKELYHNRNPRARLEFPDGVTERKQLRDKQC 422
Db 344 HVFPKQAPYARNFLQNTARAAEYVMDEYKEHFYRNPPARKEAYGDISERKLRRLRCK 403

Qy 423 KDFKWFLETYVPELHVPEDEPRGPFGLQNKGLTDYCFDYNPDENQIVGHQVILYLCHGM 482
Db 404 KSFDTYLNKVNFPNHLVPEDEPRGFWGAIARSISSECLDYNPDNNP-TGANLSLFGCHGQ 462
Qy 483 GQNOFFEYTSQKIRYNTHOPEGCIAVEAGMDTLIMHLCEE---TAPENQKFILQEDGSL 539
Db 463 GGNQFFEYTSNKSIRFNS-VTELCAEVPQKQNVGMQNCPRDGPVPANIIWHFKEDGTI 521
Qy 540 FHEQSKKCVQAAKESSDSFVPLLRDCTNSD-HQKWFPE 578
Db 522 FPHSGGLCLSAYRTPEGRPDVQ-MRTCDALDKNQIWSFEK 560

RESULT 3

US-08-648-298-2
; Sequence 2, Application US/08648298
; Patent No. 5871990
; GENERAL INFORMATION:
; APPLICANT: Henrik Clausen
; APPLICANT: Eric Paul Bennett
; TITLE OF INVENTION: UDP-N-acetyl-alpha-D-galactosamine; polypeptide
; TITLE OF INVENTION: N-acetylgalactosaminyltransferase Galnac-T3
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: NY
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/648,298
; FILING DATE: 15-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 39,475
; REFERENCE/DOCKET NUMBER: 4035/0865
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212527700
; TELEFAX: 2127536237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 633 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Submaxillary gland
US-08-648-298-2

Query Match 36.6%; Score 1142.5; DB 2; Length 633;
Best Local Similarity 44.9%; Pred. No. 1.9e-108;
Matches 244; Conservative 90; Mismatches 163; Indels 47; Gaps 14;

Qy 62 EPVMPRPVPANALGARGEAVR-LQGEELRLQESVRLHQNIVLSDRISLHRL-PE 119
Db 108 KPVLDPRPQSNAPGASGKAFKTNLSVEQKEKEGECFNAFASDRISLHRLDGLPD 167
Qy 120 RWNPLCKEKYD-YDNLPRTSVITAFYNEAWSTLRTVYVLETSFDILLEEVILVDDYS 178
Db 168 TRPECTIEQKFKRCPPLPTTSVITAFYNEAWSTLRTVSHVLSYSSPAILLKEIILVDDAS 227
Qy 179 DREHLKERLANELSLPKVRLIRANKREGLVRARLLGASARGDVLTFDCHCECHGWL 238
Db 228 VDEYHLKDLDEYKQKFSIVKIVQRERKGLITARLLGATVATETLTFDHAHCECFYGL 287

Qy 507 IAVEAGMDTLIMHLCETAPENQ-----KFILQEDGSLFHEQSKKCVQARKESSDSF 559
Db 483 LDVSKLNGFVMTLKCHH-LKGNQLWEYDPVKLTQ-----HVNQNCLDKATDE--DSQ 533
Qy 560 VPLLRDCTNSDQKWFVK 577
Db 534 VPSIRDCSGRSQWLLR 551

RESULT 7
US-08-967-508-19
; Sequence 19, Application US/08967508
; Patent No. 5910570
; GENERAL INFORMATION:
; APPLICANT: The Upjohn Company
; APPLICANT FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAC:
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual
; ADDRESSEE: Property Legal Services
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,508
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/602,830
; FILING DATE: 13 No. 5910570ember 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELEPHONE: 616-833-2210
; TELEFAX: 616-833-8897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 517 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-967-508-19

Query Match 35.2%; Score 1100; DB 2; Length 517;
Best Local Similarity 44.1%; Pred. No. 3.3e-104;
Matches 238; Conservative 84; Mismatches 168; Indels 50; Gaps 17;

Qy 56 PRGRRPVPVPPVPPANALGARGEAVRLQQLQGEELRLQBEVRLHQINIVLSRISLHR 115
Db 2 PSPA-LEPVQK---PHEGPGEMKPVVVPKEDQE--KMKEMFKINQFNLMASEWIALNR 54
Qy 116 RLPERWNPCKEKKYDYNLPRTSVIITAFYNEANSTLLRTVYVLETSPOILLBEVLVD 175
Db 55 SLDPVRLGCKTKYVP-DNLPTTSWIVFHNHNSVLLRTVHSINSPRHLSEIIVLD 113
Qy 176 DYSREHLKERLANELSLP-KVRLIRANKREGVLVRLGLGASARGDVLTLFLDCHCECH 234
Db 114 DASERDFLKPLESYVKKLVPPVHVIRWEQSGILIRLKGAAVSKGVITFLDAHCECT 173
Qy 235 EGMLEPLLQRIHEESAVVCPVIDIDWNTPEYLGNSGEPQIGGFDWRLVFTWHTVPERE 294

Db 174 VGMLEPLLARIKHDRKTVWCPFIIDVISDDTFEYMAGS-DMTYGGFNKLNFRWYVPVQRE 232
Qy 295 RIRMSQSDV-IRSPMTWAGGLFAVSKYFVYLGSDYTGMEVWGGENLEFSPRIWQCGVL 353
Db 233 MDRKKGDRTLFVRTPTWAGGLFSDRYFQBIGYDAGMDIWGGENLEISPRWQCGTL 292
Qy 354 ETHPCSHVGHVFPKQAPYS-----RNKALANSVRAAEVWMDPEFKELYHHRNPRARLE 405
Db 293 EIVTCSHVGHVFRKATPYTFPGGTQIINK---NNRRLAEVWMDPEFKNFYIISPGVTKV 349
Qy 406 PFGDVTERRKQLRDKLQCKDFKWFLETYVPELHVDPEDRPGFGLMLQNLGL-TDYCFDYNPP 464
Db 350 DYGDISSRLGLRHKLCRPFSEWYLENTYPPDSQIPRH---YFSLGEIRNVETNQCLDNMAR 406
Qy 465 DENQIVGHQVILYLCHGQGNORPEYTSQKEIRYNTHQPEGCIAVEAGMDTLIMHLCET 524
Db 407 KENEKVG---IFNCHGQGNQVFSYANKREITD-----DLCLDVSKLNGVPTWMLKCHH- 457
Qy 525 APENQ-----KFILQEDGSLFHEQSKKCVQARKESSDSFVPLLRDCTNSDQKWFVK 577
Db 458 LKGNQLWEYDPVKLTQ-----HVNQNCLDKATDE--DSQVPSIRDCSGRSQWLLR 509

RESULT 8
US-08-967-506-19
; Sequence 19, Application US/08967506
; Patent No. 6096512
; GENERAL INFORMATION:
; APPLICANT: The Upjohn Company
; APPLICANT FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAC:
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual
; ADDRESSEE: Property Legal Services
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,506
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/602,830
; FILING DATE: 13 No. 6096512ember 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELEPHONE: 616-833-2210
; TELEFAX: 616-833-8897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 517 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-967-506-19

Query Match 35.2%; Score 1100; DB 3; Length 517;
Best Local Similarity 44.1%; Pred. No. 3.3e-104;


```
Qy 48 AEPGPPTRPRGR-----EPVMPRPVPANALGA--RGEAVRLQLOGEELRLOES 97
Db 100 ALPOARRNQSGRRGGYRLIKOPRRQDKEAPKRDWGADEGEV-----SEEEETPFS 153
Qy 98 VRLHQINILSDRLSLHRLPERWNPLCKEKKYDYNLPTSVLIAPYNEAWSTLLRTVY 157
Db 154 LDPRGLOEALSARIPLQRAUPEVRHPLCLQ-QHPQDSLPTASVILCFHDEAWSTLLRTVH 212
Qy 158 SVLETSPIILLEEVILVDDYSDREHLKERLANELSGLPKVLIRANKREGVLARLILGAS 217
Db 213 SILDTVPRAFLKEIILVDDLSQQQLKSALSEYVARLEGVKLRSNKRSLAIRARMLGAT 272
Qy 218 AARGDVLTFLDCHCEHGWLEPLLQRIHEESAVVCPVIDVIDWNTFEYLGNSEGEPOIG 277
Db 273 RATGDVLVFMFDAHCECHPGWLEPLLSRIAGRSRVVSPVIDVIDWKTFOYY-PSKDLQRG 331
Qy 278 GFDWRLVFTWHTVPERIRMQSPVDVIRSPMTAGGLFAVSKYKFEYLVGSDYDTGMEVWGG 337
Db 332 VLDWKLDWFHEWPELPEHVRKALQSPISPIRSPVPGEVVAMDRHYFQNTGAYDSLMSURGG 391
Qy 338 ENLEFSFRIWQCGVLETHPCSHVGHVFPKQ---APYSRNKALANSVRAAEVMMDEFKEL 394
Db 392 ENLEFSFKAWLGGSVLEILPCSRVGHYQNDSSHPLDQEAATLRNRVRIAEATWLGSPKET 451
Qy 395 YYHRNPRA----RLEPFGDVTERKQLDKQCKDFKWFLETVYPPELVHPEDRPGFFGMLQ 450
Db 452 FYKHSPEAFSLSKAEK-PDCMERLQQLRRLCGRTFHFWLANVYPPELVSPRPSFSGKLH 510
Qy 451 NKGLTDYCFDYNPDENQIVGHQVILYLCHGMGNQOFFEYTSQKEIRYNTHQPEGCIATVE 510
Db 511 NTGL-GLCADCQA--EGDILGCPWVLAPCSDSRQQQLQHTSRKEIHFGSPQ-HLCFAVR 566
Qy 511 AGMDTLIMHLCSE--TAPENQKFIQEDGSLFHEQSKKCVQAAKSSDSFVPLLRDCTN 568
Db 567 --QEOVLQNLCTEGLAIHQHWDFOENGMIHVILSKCKMEAVVQENNKDL--YLRPCDG 622
Qy 569 SDHQKWFFKE 578
Db 623 KARQQRWFDQ 632
```

RESULT 11

```
US-09-347-488-2
; Sequence 2, Application US/09347488
; Patent No. 6239266
; GENERAL INFORMATION:
; APPLICANT: Munroe, David
; APPLICANT: Pribili, Ingrid
; TITLE OF INVENTION: ZAP-3 TUMOR ASSOCIATED GENES AND THEIR
; FILE REFERENCE: GEN-2PRV
; CURRENT APPLICATION NUMBER: US/09/347,488
; CURRENT FILING DATE: 1999-07-02
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 638
; TYPE: PRT
; ORGANISM: H. sapiens
US-09-347-488-2
```

```
Query Match 32.6%; Score 1017; DB 3; Length 638;
Best Local Similarity 42.2%; Pred. No. 1.6e-95;
Matches 232; Conservative 78; Mismatches 204; Indels 36; Gaps 14;
```

```
Qy 48 AEPGPPTRPRGR-----EPVMPRPVPANALGA--RGEAVRLQLOGEELRLOES 97
Db 100 ALPOARRNQSGRRGGYRLIKOPRRQDKEAPKRDWGADEGEV-----SEEEETPFS 153
Qy 98 VRLHQINILSDRLSLHRLPERWNPLCKEKKYDYNLPTSVLIAPYNEAWSTLLRTVY 157
Db 154 LDPRGLOEALSARIPLQRAUPEVRHPLCLQ-QHPQDSLPTASVILCFHDEAWSTLLRTVH 212
```

```
Qy 158 SVLETSPIILLEEVILVDDYSDREHLKERLANELSGLPKVLIRANKREGVLARLILGAS 217
Db 213 SILDTVPRAFLKEIILVDDLSQQQLKSALSEYVARLEGVKLRSNKRSLAIRARMLGAT 272
Qy 218 AARGDVLTFLDCHCEHGWLEPLLQRIHEESAVVCPVIDVIDWNTFEYLGNSEGEPOIG 277
Db 273 RATGDVLVFMFDAHCECHPGWLEPLLSRIAGRSRVVSPVIDVIDWKTFOYY-PSKDLQRG 331
Qy 278 GFDWRLVFTWHTVPERIRMQSPVDVIRSPMTAGGLFAVSKYKFEYLVGSDYDTGMEVWGG 337
Db 332 VLDWKLDWFHEWPELPEHVRKALQSPISPIRSPVPGEVVAMDRHYFQNTGAYDSLMSURGG 391
Qy 338 ENLEFSFRIWQCGVLETHPCSHVGHVFPKQ---APYSRNKALANSVRAAEVMMDEFKEL 394
Db 392 ENLEFSFKAWLGGSVLEILPCSRVGHYQNDSSHPLDQEAATLRNRVRIAEATWLGSPKET 451
Qy 395 YYHRNPRA----RLEPFGDVTERKQLDKQCKDFKWFLETVYPPELVHPEDRPGFFGMLQ 450
Db 452 FYKHSPEAFSLSKAEK-PDCMERLQQLRRLCGRTFHFWLANVYPPELVSPRPSFSGKLH 510
Qy 451 NKGLTDYCFDYNPDENQIVGHQVILYLCHGMGNQOFFEYTSQKEIRYNTHQPEGCIATVE 510
Db 511 NTGL-GLCADCQA--EGDILGCPWVLAPCSDSRQQQLQHTSRKEIHFGSPQ-HLCFAVR 566
Qy 511 AGMDTLIMHLCSE--TAPENQKFIQEDGSLFHEQSKKCVQAAKSSDSFVPLLRDCTN 568
Db 567 --QEOVLQNLCTEGLAIHQHWDFOENGMIHVILSKCKMEAVVQENNKDL--YLRPCDG 622
Qy 569 SDHQKWFFKE 578
Db 623 KARQQRWFDQ 632
```

RESULT 12

```
US-09-795-926-43
; Sequence 43, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 603
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-43
```

```
Query Match 31.7%; Score 991.5; DB 4; Length 603;
Best Local Similarity 36.3%; Pred. No. 6.2e-93;
Matches 231; Conservative 95; Mismatches 197; Indels 113; Gaps 19;
```



```
Qy 8 RRCPRRLRRGREALVLLALLALAGLSVLAOR-----GAGAGAAEPGPRTPRCRR 61
Db 2 RRKRLRLQVALVLAALVLLPNVGLWALYRERQDPTGGSGAIVA----- 48
Qy 62 EPVMPRPVPANALGARG-----EAVRLQLQ-----GEELR----- 92
Db 49 -----PAAGQSHSRQKKTFFLGDGQKLDKWDHKEAIRDAQRVNGEQGRPVMT 99
Qy 93 ---LQESVRLHQINILSDRISLHRLRPERWNLCKEKKYDYNLPRTSVILAFYNEAW 149
Db 100 DAERVQDAYRENGFNIVYSDKISLNSRLPDTRHPNCNKRY-LETLPNTSIIPFHNEGW 158
Qy 150 STLARTVYVLETSPIILBEVILVDDYSDEHLKERLANELSGLPKVLIRANKREGVL 209
Db 159 SSLRTVHSLNRPPELVAEIVLVDPSDEHLKPLEDMALFSPVRIILTKKREGLI 218
Qy 210 RARLLGASARGDVLTFDCHCEHGLEPLQRIHEESAVVCPVIDIDMNTFFYLG 269
Db 219 RTRMLGASVATGDTVITFLDSCHEANVNWLPPLLDRIARNRKTIVCPMIDVIDHDDFRYET 278
Qy 270 NSGEPQGGFWRLVFTWHTVPERIRMQSPVDVIRSPMTAGGLFAVSKKYFYLGSDY 329
Db 279 QAGDAMRGAFDWEWY--KRIPPELQKADPSDPFSPVMAGGLFAVDRKWFELGGYD 336
Qy 330 TGMVWGGENLEFSPRIWOCGGVLETHPCSHVGHVFPKQAPYS--RNKALA-NSVRAAEV 386
Db 337 PGLSEIWGEQVEISFKVMWCGGRMEDIPCSRVGHIYRKYVPYKVPAGVSLARNLKRVAEV 396
Qy 387 WMDSPFKELYHRNPRARLEPPGDVTERKQLRDLQCKDFKWFLETV---YPELHVPEDRP 443
Db 397 WMDVAYEVIYQRRPEYRHLNLSAGDAVQKLSLNSCKSFKFMFKIAWDLPKFYPPVEPP 456
Qy 444 -GFGMLQNGKLTDCFPYPPDENQIVGHQVILYLC-HGMQO-----NQFFEYTSQKEI 496
Db 457 AAANGEIRNVG-TGLCAD----TKHGALGSLRLEGCVGRGEAAWNMQVFTTREDI 511
Qy 497 RYN--THQPEGIAVEAGMDTLIMHLCETAPENQKFIQEDGSLFHEQSKKCVQAAARKE 554
Db 512 RPDGPQHTKFCFPAISHTSPVTLYDC-HSMKGNQLWKYRKDKTLIHPVSGSCM----- 564
Qy 555 SSSDFVPLLRDCTNSDH-----QKWFKE 577
Db 565 -----DCSESDHRIFMNTCNPSLTQOQWLFE 590

RESULT 13
US-09-795-926-31
; Sequence 31, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kiege, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849

PRIORITY FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 506
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-31

Query Match 31.2%; Score 974.5; DB 4; Length 506;
Best Local Similarity 40.3%; Pred. No. 2.6e-91;
Matches 207; Conservative 84; Mismatches 167; Indels 53; Gaps 14;

Qy 95 EESVRLHQINILSDRISLHRLRPERWNLCKEKKYDYNLPRTSVILAFYNEAWSTLLR 154
Db 8 DQAYRENGFNIVYSDKISLNSRLPDTRHPNCNKRY-LETLPNTSIIPFHNEGWSSLLR 66
Qy 155 TVYSVLETSPIILBEVILVDDYSDEHLKERLANELSGLPKVLIRANKREGVLRL 214
Db 67 TVHSLNRPPELVAEIVLVDPSDEHLKPLEDMALFSPVRIILTKKREGILTRML 126
Qy 215 GASAARGDVLTFDCHCEHGLEPLQRIHEESAVVCPVIDIDMNTFFYLGNSGEP 274
Db 127 GASVATGDTVITFLDSCHEANVNWLPPLLDRIARNRKTIVCPMIDVIDHDDFRYETQAGDA 186
Qy 275 QIGGFWRLVFTWHTVPERIRMQSPVDVIRSPMTAGGLFAVSKKYFYLGSDYDTGMEV 334
Db 187 MRGAFDWEWY--KRIPPELQKADPSDPFSPVMAGGLFAVDRKWFELGGYDPLGLEI 244
Qy 335 WGGENLEFSPRIWOCGGVLETHPCSHVGHVFPKQAPYS--RNKALA-NSVRAAEVWMDDEF 391
Db 245 WGGQVEISFKVMWCGGRMEDIPCSRVGHIYRKYVPYKVPAGVSLARNLKRVAEVMWDEY 304
Qy 392 KELYHRNPRARLEPPGDVTERKQLRDLQCKDFKWFLETV---YPELHVPEDRP-CFFG 447
Db 305 AEYIYQRRPEYRHLNLSAGDAVQKLSLNSCKSFKFMFKIAWDLPKFYPPVEPPAAAG 364
Qy 448 MLQNGKLTDCFPYPPDENQIVGHQVILYLC-HGMQO-----NQFFEYTSQKEIRYN-- 499
Db 365 EIRNVG-TGLCAD----TKHGALGSLRLEGCVGRGEAAWNMQVFTTREDIRPGDP 419
Qy 500 THQPEGIAVEAGMDTLIMHLCETAPENQKFIQEDGSLFHEQSKKCVQAAARKESSDSF 559
Db 420 QHTKKFCFPAISHTSPVTLYDC-HSMKGNQLWKYRKDKTLIHPVSGSCM----- 467
Qy 560 VPLLRDCTNSDH-----QKWFKE 577
Db 468 -----DCSESDHRIFMNTCNPSLTQOQWLFE 493

RESULT 14
US-09-795-926-41
; Sequence 41, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kiege, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US/09795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
```

```
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 631
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-41

Query Match      31.0%; Score 967; DB 4; Length 631;
Best Local Similarity 34.3%; Pred. No. 2,28-90;
Matches 231; Conservative 95; Mismatches 197; Indels 142; Gaps 20;

QY      8 RRCPRRLRRGREALVLLALLALAGLGLSVLRAQR-----GAGAGAAEPGPPRTPRGRR 61
DB      2 RRREKRLQVALVLAALVLLPNVGLMALYRERQDGTGGGAAVA----- 48
QY      62 EPVMPRPVPANALGARG-----EAVRLQLQ-----GEELR----- 92
DB      49 -----PAAGGSHSRQKTFPLGDGQKLDKDHKEAIRRDAQRVGNAGEQRPYPMT 99
QY      93 ---LQESVRLHQINIVLSDRISLHRRLPERNWPLCKEKKYDYNLPRTSVITAFYNEAW 149
DB     100 DAERVDQAYRENGNIYVSDKISLNSRLPDTRHPCNCKRY-LETLPTSIIIPFHNEGW 158
QY     150 STLLRTVYSLVETSPDILLLEEVILVDDYSREHLKERLANELSLGLPKVRLIRANKRGVL 209
DB     159 SLLRLTVHSLNRSPPLEVAEIVLDDFSREHLKPKLEDMYALFPFVSRILRTKREGLI 218
QY     210 RARLGASAAAGDVLTFDCHCEHEGWLEPLLQRIHEESAVVCPVIDVIDMNTFYLG 269
DB     219 RTRMLGASVATGDIITLDSHCEANVNLPLLDRIARNRKTIVCPMIDVIDHDDFRYET 278
QY     270 NSEGPQIGFDWRLVFTWHTVPERERIRMSQSPDVIRSPMTWAGGLFAYSKKYFEYLGSGYD 329
DB     279 QAGDAMRGAFDWMY--KRIPPELOKADSPDPFESVWAGGLFAVDRKWFELGGYD 336
QY     330 TGMVWNGENLEFSR-----IWQCGVLETHPCSH 360
DB     337 PGLIEWGEQVEISFKGLHMLPRLVSNWPQAVFLPRAPNMLAQVWMCGRMEDIPCSR 396
QY     361 VGHVFPKQAPYS--RNKALA-NSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTTERKOLR 417
DB     397 VGHYRYKYVYKYPAGVSLARNLKRVAEVMWDEYAEYIYQRRPYRHLISAGDVAQKKLR 456
QY     418 DKLQCKDFKWFLETV---YPELHVPEDRP-GFFGMLQNKGLTDYCFDYNPPDENQIVGHQ 473
DB     457 SSLNCKSFKNWMTKIAMDLPKFYPPVPPAAWGEIRNVG-TGLCAD---TKHGALGSP 511
QY     474 VILYLC-HGMGQ-----NOFEYTSQKEIRYN--THOPEGCIAVEAGMDTLIMHLCETA 525
DB     512 LRLEGCVRGGEAAWNNQVFTFWREDIRPGDPQHTKKCFDAISHTSPVTLDYC-HSM 570
QY     526 PENQKFIQDGSFLFHEQSKKCVQAARKSSDSFVPLLRDCTNSDH-----Q 572
DB     571 KGNQLWKYRKDTLYHPVSGSCM-----DCSESDDRIFMNTCNPSLTLQ 614
QY     573 KNPKF 577
DB     615 QWLFE 619
```

RESULT 15

```
US-09-795-926-29
; Sequence 29, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Halbun, Brin
```

```
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 535
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-29
```

Query Match 30.4%; Score 950; DB 4; Length 535;

Best Local Similarity 38.3%; Pred. No. 9,4e-89;

Matches 207; Conservative 84; Mismatches 167; Indels 82; Gaps 15;

```
QY      95 EESVRLHQINIVLSDRISLHRRLPERNWPLCKEKKYDYNLPRTSVITAFYNEAWSTLLR 154
DB      8 DQAYRENGNIYVSDKISLNSRLPDTRHPCNCKRY-LETLPTSIIIPFHNEGWSELLR 66
QY     155 TVYSVLETSPDILLLEEVILVDDYSREHLKERLANELSLGLPKVRLIRANKRGVLRLRL 214
DB     167 TVHSLNRSPPLEVAEIVLDDFSREHLKPKLEDMYALFPFVSRILRTKREGLI RTML 126
QY     215 GASAAAGDVLTFDCHCEHEGWLEPLLQRIHEESAVVCPVIDVIDMNTFYLGSGYD 274
DB     127 GASVATGDIITLDSHCEANVNLPLLDRIARNRKTIVCPMIDVIDHDDFRYETQAGDA 186
QY     275 QIGGFDRVLTWHTVPERERIRMSQSPDVIRSPMTWAGGLFAYSKKYFEYLGSDYTGMEV 334
DB     187 MRGAFDWMY--KRIPPELOKADSPDPFESVWAGGLFAVDRKWFELGGYDPLGLEI 244
QY     335 WGGENLEFSR-----IWQCGVLETHPCSHVGHVF 365
DB     245 WGGEQVEISFKGLHMLPRLVSNWPQAVFLPRAPNMLAQVWMCGRMEDIPCSRUGHYI 304
QY     366 PKQAPYS--RNKALA-NSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTTERKOLRDKQC 422
DB     305 RKVVPKYVYKYPAGVSLARNLKRVAEVMWDEYAEYIYQRRPYRHLISAGDVAQKKLSLNC 364
QY     423 KDFKWFLETV---YPELHVPEDRP-GFFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYL 478
DB     365 KSKFWMTKIAMDLPKFYPPVPPAAWGEIRNVG-TGLCAD---TKHGALGSPRLLEG 419
QY     479 C-HGMGQ-----NOFEYTSQKEIRYN--THOPEGCIAVEAGMDTLIMHLCETAPENQK 530
DB     420 CVRGGEAAWNNQVFTFWREDIRPGDPQHTKKCFDAISHTSPVTLDYC-HSMKGNQL 478
QY     531 FILQEDGSLFHEQSKKCVQAARKSSDSFVPLLRDCTNSDH-----QKWF 577
DB     479 WKYRKDTLYHPVSGSCM-----DCSESDDRIFMNTCNPSLTLQWLFE 522
```

Search completed: December 4, 2003, 19:44:49

Job time : 23 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 19:43:48 ; Search time 36 Seconds
(without alignments)
3001.570 Million cell updates/sec

Title: US-10-074-527-2

Perfect score: 3124

Sequence: 1 MWGTTARRRCPRELRRGREA.....LLRDCTNSDQKWFKERM 581

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3124	100.0	581	14	US-10-074-527-2
2	3076	98.5	581	12	Sequence 2, Appli
3	1675	53.6	578	14	US-10-292-896-58
4	1369	43.8	256	9	US-10-074-527-8
5	1369	43.8	256	11	US-09-925-299-992
6	1309	41.9	316	15	US-09-925-299-992
7	1304	41.7	289	11	US-10-106-698-4875
8	1176	37.6	626	14	US-09-895-298-95
9	1142.5	36.6	633	15	US-10-001-851-27
10	1122.5	35.9	556	12	US-10-205-823-146
11	1109	35.5	559	14	US-10-292-896-62
12	1101.5	35.3	559	14	US-10-001-851-22
13	1101.5	35.3	559	14	US-10-001-851-20
14	1100.5	35.2	551	9	US-09-925-301-1006
15	1097	35.1	559	12	US-10-001-851-23
			559	14	US-10-205-219-36

16	1097	35.1	559	12	US-10-205-219-76	Sequence 76, Appl
17	1097	35.1	559	14	US-10-001-851-21	Sequence 21, Appl
18	1097	35.1	559	14	US-10-001-851-24	Sequence 24, Appl
19	1023	32.7	639	9	US-09-789-417-2	Sequence 2, Appli
20	1023	32.7	639	9	US-09-792-451-2	Sequence 2, Appli
21	1023	32.7	639	11	US-09-946-374-347	Sequence 347, App
22	1023	32.7	639	12	US-10-015-387A-347	Sequence 347, App
23	1023	32.7	639	12	US-10-006-130A-347	Sequence 347, App
24	1023	32.7	639	12	US-10-199-672-364	Sequence 347, App
25	1023	32.7	639	12	US-10-006-172A-347	Sequence 347, App
26	1023	32.7	639	12	US-10-187-749-364	Sequence 364, App
27	1023	32.7	639	12	US-10-194-457-364	Sequence 364, App
28	1023	32.7	639	12	US-10-184-642-364	Sequence 364, App
29	1023	32.7	639	12	US-10-196-747-364	Sequence 364, App
30	1023	32.7	639	12	US-10-015-392A-347	Sequence 347, App
31	1023	32.7	639	12	US-10-017-253A-347	Sequence 347, App
32	1023	32.7	639	12	US-10-173-689-364	Sequence 364, App
33	1023	32.7	639	12	US-10-173-690-364	Sequence 364, App
34	1023	32.7	639	12	US-10-173-691-364	Sequence 364, App
35	1023	32.7	639	12	US-10-173-692-364	Sequence 364, App
36	1023	32.7	639	12	US-10-173-694-364	Sequence 364, App
37	1023	32.7	639	12	US-10-173-698-364	Sequence 364, App
38	1023	32.7	639	12	US-10-173-699-364	Sequence 364, App
39	1023	32.7	639	12	US-10-173-707-364	Sequence 364, App
40	1023	32.7	639	12	US-10-174-569-364	Sequence 364, App
41	1023	32.7	639	12	US-10-174-583-364	Sequence 364, App
42	1023	32.7	639	12	US-10-174-587-364	Sequence 364, App
43	1023	32.7	639	12	US-10-174-589-364	Sequence 364, App
44	1023	32.7	639	12	US-10-174-591-364	Sequence 364, App
45	1023	32.7	639	12	US-10-175-736-364	Sequence 364, App

ALIGNMENTS

RESULT 1

US-10-074-527-2
; Sequence 2, Application US/10074527
; Publication No. US20020142426A1
; GENERAL INFORMATION:
; APPLICANT: Olandt, Peter J.
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Galvin, Katherine A.
; APPLICANT: Millennium Pharmaceuticals Inc.
; TITLE OF INVENTION: 33945, A Human Glycosyltransferase and
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: MPI2001-018P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/074,527
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/269202
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 581
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-074-527-2

Query Match	100.0%;	Score 3124;	DB 14;	Length 581;
Best Local Similarity	100.0%;	Pred. NO. 7.8e-283;		
Matches 581;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MWGTARRRCPRELRRGREALLLALLAGLSVLRAGRGAGAGAEPCPPPTPPGR	60	
Db	1	MWGTARRRCPRELRRGREALLLALLAGLSVLRAGRGAGAGAEPCPPPTPPGR	60	
Qy	61	REPVPMPPPANALGARGAVRLQGEELRQESVRLHQINYLSDRLSHRLRPER	120	
Db	61	REPVPMPPPANALGARGAVRLQGEELRQESVRLHQINYLSDRLSHRLRPER	120	
Qy	121	WNPICKEKKYDYNLPRTSVIIAFYNEAWSTLLRTVYVLETSPDILLEVLVDDYSDR	180	

Db 121 WNPLCKEKKYDYNLPTSVIIAFYNEAWSTLLRTVSVLETSPIILLEEVILVDDYSDR 180
Qy 181 EHLKERLANELSGLPKVRLLIRANKREGVRLVRLGASAGDVLTFDCHCECHGWLEP 240
Db 181 EHLKERLANELSGLPKVRLLIRANKREGVRLVRLGASAGDVLTFDCHCECHGWLEP 240
Qy 241 LLQRIHEESAVVCPVIDVDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERIRMQS 300
Db 241 LLQRIHEESAVVCPVIDVDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERIRMQS 300
Qy 301 PVDVIRPTWAGGLFAVSKYFYLYGSDYTGMEVWNGENLEFSFRIWCCGVLETHPCSH 360
Db 301 PVDVIRPTWAGGLFAVSKYFYLYGSDYTGMEVWNGENLEFSFRIWCCGVLETHPCSH 360
Qy 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTERRKQDRKL 420
Db 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTERRKQDRKL 420
Qy 421 QCKDFKWFLETVPELVHPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
Db 421 QCKDFKWFLETVPELVHPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
Qy 481 GMGQNOFFEYTSQKEIRYNTHQPEGCIATVAGMDTLIMHLCETAPENQKFIQEDGSLF 540
Db 481 GMGQNOFFEYTSQKEIRYNTHQPEGCIATVAGMDTLIMHLCETAPENQKFIQEDGSLF 540
Qy 541 HEQSKKCVQAARKESDVSFVPLLRDCTNSDHQKWFKERM 581
Db 541 HEQSKKCVQAARKESDVSFVPLLRDCTNSDHQKWFKERM 581

RESULT 2

US-10-292-896-58

; Sequence 58, Application US/10292896

; Publication No. US20030186850A1

; GENERAL INFORMATION:

; APPLICANT: HASSAN, Helle

; APPLICANT: REIS, Celso A.

; APPLICANT: BENNETT, Eric P.

; APPLICANT: CLAUSEN, Henrik

; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GA

; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS

; TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICAMENTS

; FILE REFERENCE: 4305/1H154-US3

; CURRENT APPLICATION NUMBER: US/10/292,896

; CURRENT FILING DATE: 2003-03-17

; PRIOR APPLICATION NUMBER: US 60/425,204

; PRIOR FILING DATE: 2002-11-08

; PRIOR APPLICATION NUMBER: PCT/DK01/00328

; PRIOR FILING DATE: 2001-05-10

; PRIOR APPLICATION NUMBER: US 60/203,331

; PRIOR FILING DATE: 2000-05-11

; NUMBER OF SEQ ID NOS: 130

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 58

; LENGTH: 581

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-10-292-896-58

Query Match 98.5%; Score 3076; DB 12; Length 581;

Best Local Similarity 98.5%; Pred. No. 2.4e-278;

Matches 572; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MWGTARRRCPRELRRGREALLVLLALAGLSVLRQAQAGAGAAEPGPPTPRGR 60
Db 1 MWGTARRRCPRELRRGREALLVLLALAGLSVLRQAQAGAGAAEPGPPTPRGR 60
Qy 61 REPWMPRPVPPANALGARGEAVRLQOGEELRQESVRLHQINYLSDRSLHRLRPER 120
Db 61 REPWMPRPVPPANALGARGEAVRLQOGEELRQESVRLHQINYLSDRSLHRLRPER 120
Qy 121 WNPLCKEKKYDYNLPTSVIIAFYNEAWSTLLRTVSVLETSPIILLEEVILVDDYSDR 180

Db 121 WNPLCKEKKYDYNLPTSVIIAFYNEAWSTLLRTVSVLETSPIILLEEVILVDDYSDR 180
Qy 181 EHLKERLANELSGLPKVRLLIRANKREGVRLVRLGASAGDVLTFDCHCECHGWLEP 240
Db 181 EHLKERLANELSGLPKVRLLIRANKREGVRLVRLGASAGDVLTFDCHCECHGWLEP 240
Qy 241 LLQRIHEESAVVCPVIDVDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERIRMQS 300
Db 241 LLQRIHEESAVVCPVIDVDWNTFEYLGNSGEPQIGGDFWRLVFTWHTVPERIRMQS 300
Qy 301 PVDVIRPTWAGGLFAVSKYFYLYGSDYTGMEVWNGENLEFSFRIWCCGVLETHPCSH 360
Db 301 PVDVIRPTWAGGLFAVSKYFYLYGSDYTGMEVWNGENLEFSFRIWCCGVLETHPCSH 360
Qy 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTERRKQDRKL 420
Db 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHNRPRARLEPFGDVTERRKQDRKL 420
Qy 421 QCKDFKWFLETVPELVHPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
Db 421 QCKDFKWFLETVPELVHPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
Qy 481 GMGQNOFFEYTSQKEIRYNTHQPEGCIATVAGMDTLIMHLCETAPENQKFIQEDGSLF 540
Db 481 GMGQNOFFEYTSQKEIRYNTHQPEGCIATVAGMDTLIMHLCETAPENQKFIQEDGSLF 540
Qy 541 HEQSKKCVQAARKESDVSFVPLLRDCTNSDHQKWFKERM 581
Db 541 HEQSKKCVQAARKESDVSFVPLLRDCTNSDHQKWFKERM 581

RESULT 3

US-10-074-527-8

; Sequence 8, Application US/10074527

; Publication No. US20020142426A1

; GENERAL INFORMATION:

; APPLICANT: Olandt, Peter J.

; APPLICANT: Meyers, Rachel E.

; APPLICANT: Galvin, Katherine A.

; APPLICANT: Millennium Pharmaceuticals Inc.

; TITLE OF INVENTION: 33945, A Human Glycosyltransferase and

; TITLE OF INVENTION: Uses Therefor

; FILE REFERENCE: MP12001-018P1RCP1(M)

; CURRENT APPLICATION NUMBER: US/10/074,527

; CURRENT FILING DATE: 2002-02-12

; PRIOR APPLICATION NUMBER: 60/269202

; PRIOR FILING DATE: 2001-02-15

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: Fast-Seq for Windows Version 4.0

; SEQ ID NO 8

; LENGTH: 578

; TYPE: PRT

; ORGANISM: Mus musculus

; US-10-074-527-8

Query Match 53.6%; Score 1675; DB 14; Length 578;

Best Local Similarity 57.9%; Pred. No. 2e-147;

Matches 330; Conservative 78; Mismatches 144; Indels 18; Gaps 8;

Qy 23 VLLALLALAGL-----GSVLRQAQAGAGAAEPGPPTPRGR-----PVMPRPVPA 72
Db 13 LLLALLTLAVLVEFVSSTLYASFGAG-GARELGPRLPDLDTREEDLSQPLYIKPADS 71
Qy 73 NALGARGEAVRLQOGEELRQESVRLHQINYLSDRSLHRLRPERWNPCKEKKYD 132
Db 72 HALGWEGRASKLQNEGELKQSELIBRYAINYLSDRSLHRRHIEDKRMVECKAKPHY 131
Qy 133 DNLPTSVIIAFYNEAWSTLLRTVSVLETSPIILLEEVILVDDYSDRHLKELANELS 192
Db 132 RSLPTTSVIIAFYNEAWSTLLRTVSVLETSPIILLEEVILVDDYSDRHLKELANELS 191
Qy 193 GLPKVRLIRANKREGVRLVRLGASAGDVLTFDCHCECHGWLEPPLQRIHEESAV 252

Db 192 NLERVLRIRNKREGLVRLRIGATFATGDLVLFDCHECNCWLEPLLERISRDETAI 251
Qy 253 VCPVIDVNDWNTPEYLGNSGEPQIGGDFWRLVFTWHTVPERIRMQSPVDVIRSPTMAG 312
Db 252 VCPVIDVNDWNTPEYLGNSGEPQIGGDFWRLVFTWHTVPERIRMQSPVDVIRSPTMAG 311
Qy 313 GLFAVSKKYFEYLGSDYTGMEVWNGENLEFSFRWQCGGVLETHPCSHVGHVFPKQAPYS 372
Db 312 GLFAVSKKYFOYLGSDYTGMEVWNGENLEFSFRWQCGGVLETHPCSHVGHVFPKQAPYA 371
Qy 373 RNKALANSVRAAEVWMDPEFKELYHYRNPRLRLEPFGDVTERKQLRDKLOCKDFKWFLETV 432
Db 372 RPNFLQNTAAAEVWMDPEFKELYHYRNPRLRLEPFGDVTERKQLRDKLOCKDFKWFLETV 431
Qy 433 YPELVHPEDRPGFGLMQLKGLTDYCFDYNPPDENQIVGHQVILYLCHGMGQNGOFFEYTS 492
Db 432 FSNLHVPEDRPGFGLMQLKGLTDYCFDYNPPDENQIVGHQVILYLCHGMGQNGOFFEYTS 490
Qy 493 QKEIRYNTHOPEGCIAYEAGHDTLMHLCEE---TAPENQKFILOEDGSLFHEOSKCVQ 549
Db 491 NKEIRFNS-VTELCAEVPQKDYVGMQNCPCDGLPVPVNIWHFKEDGTIFHPHTRLCLS 549
Qy 550 AAKESSDSFVPLLRDCTNSD-HQKWFPE 578
Db 550 AYTAEGRPSVH-MKTCDAIDKQKWLRF 578

RESULT 4

US-09-925-299-992
; Sequence 992, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 992
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (229)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-992

Query Match 43.8%; Score 1369; DB 9; Length 256;
Best Local Similarity 99.6%; Pred. No. 2.3e-119;
Matches 247; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 238 LEPLQRIHEESAVVCPVIDVNDWNTPEYLGNSGEPQIGGDFWRLVFTWHTVPERIR 297
Db 6 LEPLQRIHEESAVVCPVIDVNDWNTPEYLGNSGEPQIGGDFWRLVFTWHTVPERIR 65
Qy 298 MQSPVDVIRSPTMAGGLFAVSKKYFEYLGSDYTGMEVWNGENLEFSFRWQCGGVLETHP 357
Db 66 MQSPVDVIRSPTMAGGLFAVSKKYFEYLGSDYTGMEVWNGENLEFSFRWQCGGVLETHP 135
Qy 358 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHYRNPRLRLEPFGDVTERKQLR 417
Db 126 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHYRNPRLRLEPFGDVTERKQLR 185
Qy 418 DKLOCKDFKWFLETVPELVHPEDRPGFGLMQLKGLTDYCFDYNPPDENQIVGHQVILY 477
Db 186 DKLOCKDFKWFLETVPELVHPEDRPGFGLMQLKGLTDYCFDYNPPDENQIVGHQVILY 245

RESULT 6

US-10-106-698-4875
; Sequence 4875, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03

Qy 478 LCHGMGQN 485
Db 246 LCHGMGQN 253

RESULT 5

US-09-925-299-992
; Sequence 992, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 992
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (229)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-992

Query Match 43.8%; Score 1369; DB 11; Length 256;
Best Local Similarity 99.6%; Pred. No. 2.3e-119;
Matches 247; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 238 LEPLQRIHEESAVVCPVIDVNDWNTPEYLGNSGEPQIGGDFWRLVFTWHTVPERIR 297
Db 6 LEPLQRIHEESAVVCPVIDVNDWNTPEYLGNSGEPQIGGDFWRLVFTWHTVPERIR 65
Qy 298 MQSPVDVIRSPTMAGGLFAVSKKYFEYLGSDYTGMEVWNGENLEFSFRWQCGGVLETHP 357
Db 66 MQSPVDVIRSPTMAGGLFAVSKKYFEYLGSDYTGMEVWNGENLEFSFRWQCGGVLETHP 125
Qy 358 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHYRNPRLRLEPFGDVTERKQLR 417
Db 126 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHYRNPRLRLEPFGDVTERKQLR 185
Qy 418 DKLOCKDFKWFLETVPELVHPEDRPGFGLMQLKGLTDYCFDYNPPDENQIVGHQVILY 477
Db 186 DKLOCKDFKWFLETVPELVHPEDRPGFGLMQLKGLTDYCFDYNPPDENQIVGHQVILY 245
Qy 478 LCHGMGQN 485
Db 246 LCHGMGQN 253

Query Match	36.6%;	Score 1142.5;	DB 15;	Length 633;
Best Local Similarity	44.9%;	Pred. No. 1.3e-97;		
Matches 244;	Conservative	90;	Mismatches 163;	Indels 47; Gaps 14;
Qy	62	EPVMPRPVPANALGARGEAVR-LQLQGEELRLQEESVRLHQINIVLSDRISLHRLR-PE	119	
Db	108	KPVLDPRQPDSNAPGASGKAFKTNLSVEQEKERGEAKHCFNAPASDRISLHRLDLPD	167	
Qy	120	RWNPLCKEKKYD-YDNLPRTSVIIAFYNEAWSTLLRRTVYSVLETSPPDILLEEIVLDDYS	178	
Db	168	TRDPECTEQKFKRCPPLPTTSIIIVFNEAWSTLLRRTVHVSVLVSSPAILLKEIILVDDAS	227	
Qy	179	DREHLKERLANEISGLPKVRLIRANKREGIVARLLGLASAARGDVLTFLOCHCEHEGWL	238	
Db	228	VDEYLDHLDDEYVQKFSIVIKVIRORERKGLITAKLLGATVATATLLTFLDHAECFCYGWL	287	
Qy	239	EPLLRIHEBESAVCVSDIVIDWNTFEYILGNSGP-----QIGGFQDWRLVFTWHTVPER	293	
Db	288	EPLLARIAENVTAVVSPIASIDLNTPEF--NKSPSPYGSNNHNGFNDWSLSFGWESUPDH	345	
Qy	294	ERTMQSPDVIIRSPHTWAGGLFAVSKKYFEYLSYDTCMEVWGGENLEFSPFRWQCGGVL	353	
Db	346	EKORRKDETYPKITPTTAGGLFSISKSEFYIGSYDEEMEWGGENIENSFVWQCGGQL	405	
Qy	354	ETHPSCSHGVHVPFKQAPYSRNKA----LANSVRAAEVWMDSEFKELYHRRNPRA-----RLE	405	
Db	406	EINPCSVVGHVFRSKSPHSPKPGTQVTRNQVRLAEVWMDSEYKEIFYRNTDAAKIVQKQ	465	
Qy	406	PFGDVTERRQLRDKLQCKDPKFWLFETVYVPELHVDPEDRPGFGMLQNKGLTDTYCFDYNPPD	465	
Db	466	AFGDSLKRFRFKHRLCKNTFWYLNNTYPEYVVDLNPVLSGIKSYGV-OPFLCIDVG---S21		

Query Match	35.9%;	Score 1122.5;	DB 12;	Length 556;
Best Local Similarity	46.3%;	Pred. No. 8e-96;		
Matches	239;	Conservative	78;	Mismatches 170;
				Indels 29;
				Gaps 12
Qy	76	GARCEAVRLQLQGCELRLQRESVRLHGINVYLSDRISLHRLPRRNWPLCKEKKYVDYDNL	135	
Db	58	GEMGKAVLIIPKDDG--KMKELFKINGFNLMASDLIALNRLSLPDRVLECGCKTKVP-DEL	114	
Qy	136	PTTSVIITAFYNEAWSTLLRTTVYSYLETSPOILLBEVILVDDYSDREHLKERLANELSGLP	195	
Db	115	PNTSWIIVFHNEAWSTLLRTTVYSVINRSPHYLLSEVILVDDASERDFLKLTLNLYVKNLE	174	
Qy	196	-KVLIRANKREGIIVARLILCASARGDVLTFDLCHCEHGWLEPLLORIHBEESAVVC	254	
Db	175	VPVKIIIRBEBSGLIIRARLGAASKGVITFLDAHCETCTGWLLEPLIARIKEDRTVVC	234	
Qy	255	PVIDVIDWNTFEYILGNSGEPQIGFDRWLTVFTWHTVPERERIRMQSPVDV-IRSPTMAGG	313	
Db	235	PIIDVISDDITFEYMAGS-DMTYGGFNKMLNFRWYVPVQREMDRRKGDRTLPVPTPTMAGG	293	
Qy	314	LPAVSKYKFEYLSYDTGMEVWGGENILEFSFRIWQCGGVLETHPCSHVGHVFPKQAPYS-	372	
Db	294	LFSIDRNYFEIGTYDAGMDIWMGGENLEMSFRIWQCGGSLBIVTCSHVGHVFRKATPYTF	353	
Qy	373	-----RNKALANSVRAAEVWMDDEFKELYVHRNPRARLBPFVDTERKQLRDKLOCKDF	425	
Db	354	PGGTGHVINK---NNRLAEVWMDDEFKDFVFIISPGVVKVVDYGDVSVRKTIRENLKCRPF	410	
Qy	426	KWFLFTVYVPELHVPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCHGMGON	485	
Db	411	SWYLENIYPPSOIIPR-RYYSLSIGLRNVE-TNOCIDNNRKNENEKVG----FNCHGMGON	464	

Qy	486	QFFETYSQKEIRYNTHQPEGCIAPAGMDTILMHLCEETAPENQKFILQDDGSLFHEQSK	545
Db	465	QFYSYTADEKIRTD----DLCLDVSRLNGPVMILKCHHMRGNQLWEYDAERLTLRHVN	520
Qy	546	KCVQAARKSSSFVPLLRDCTNSDHQKWFPERML	581
Db	521	OCLDEPSES--DKWVFTMODCGSRSOOWLLRNMTL	554

RESULT 11
 US-10-001-851-22
 ; Sequence 22, Application US/10001851
 ; Publication No. US20020115628A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MEYERS, Rachel A.
 ; APPLICANT: WILLIAMSON, Mark
 ; TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1el Human Glycosyl Transferase
 ; TITLE OF INVENTION: Uses Thereof
 ; FILE REFERENCE: 10147-5601
 ; CURRENT APPLICATION NUMBER: US/10/001.851
 ; CURRENT FILING DATE: 2001-11-20
 ; PRIOR APPLICATION NUMBER: US 60/249,939
 ; PRIOR FILING DATE: 2000-11-20
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 22
 ; LENGTH: 559
 ; TYPE: PRT
 ; ORGANISM: Mus sp.
 US-10-001-851-22

Db 513 LTIQ-----HVNNSNOCLDKATEE--DSQVPSIRDCTGSRSQOWLLR 551

RESULT 12

US-10-001-851-20

; Sequence 20, Application US/10001851

; Publication No. US20020115628A1

; GENERAL INFORMATION:

; APPLICANT: MEYERS, Rachel A.

; APPLICANT: WILLIAMSON, Mark

; TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1el Human Glycosyl Tr

; TITLE OF INVENTION: Uses thereof

; FILE REFERENCE: 10147-56U1

; CURRENT APPLICATION NUMBER: US/10/001,851

; CURRENT FILING DATE: 2001-11-20

; PRIOR APPLICATION NUMBER: US 60/249,939

; PRIOR FILING DATE: 2000-11-20

; NUMBER OF SEQ ID NOS: 29

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 20

; LENGTH: 559

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-001-851-20

RESULT 13
US-09-925-301-1006

```

; SEQ ID NO 1006
;
; LENGTH: 561
;
; TYPE: PRT
;
; ORGANISM: Homo sapiens
US-09-925-301-1006

```

```

Query Match      35.2%; Score 1100.5; DB 14; Length 559;
Best Local Similarity 43.0%; Pred. No. 9.2e-94;
Matches 240; Conservative 86; Mismatches 171; Indels 61; Gaps 17;

Qy 38 RAORGAGAGAAEPGPRTPRGRPRPMPRPVPANALGARGEAVRLQLOGEELRLQOES 97
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 37 KXERGLPAGDV-----LEPVQK-----PHEGPGEMKPVVIPKEDQE--KXKEM 78
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 98 VRUHQINIVLSDRISUHRRLPERMNPCLCKEKKYDYNLPRTSVIIAFYNBAWSTLLRTVY 157
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 79 FKINQFNLMASEMIALNRLSPDVRLEGCKTKVP-DNLPTTSVVIVPHNEAWSTLLRTVH 137
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 158 SVLETPSDILLERVIIVDDYSREHLKERLANELSLGP-KVRLIRANKRGLVYARLLGA 216
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 138 SVINRPRHLEBEIVLVDASEHDFLKRPLESVYKKLKVPHVIRMEQRSGLIRARLKGA 197
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 217 SAARGDVLTFDCHCECHGWLPEPLLQRIHEESAVVPCVIDIDWNTFEYVLGNSGSPQI 276
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 198 AVSGQVITFLDAHCECTGVGLEPLLARIKHDKRTVVCPIIDVISDTEFYMAGS-DMTY 256
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 277 GGFDRVLVFTWHTVPERIRIMQSPVDV-IRSTNAGGLPAVSKKPYEYLGSDYDTGMENV 335
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 257 GGFNWKLNFRWYVPVQEMDRRKGRDRTLPRVPTNAGGLFSIDRDYFQEITGYDAGMDIW 316
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 336 GGENLEFSFRIWOCGGVLETHPCSHGVHPFKQAPYS-----RNKALANSVRAAEVW 387
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 317 GGENLEISFRIWOCGGTLEIVTCSHGVHFRKATPYTFPGTGQIINK---NNRRLAEVW 373
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 388 MDEFKELYHRNPRARLEPFGDVTERQRLDKLOCKDFKWFLETVYPELVHPEDRPGFFG 447
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 374 MDEFKNFFYIISPGVTKVDYGDIISSRLGLRHLKQCRPFSWYLENIYVDSQIPRH---YFS 430
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 448 MLQNKGL-ITDCYDYNPPDENQIVGHQVILYLCHGQGNQFFYTSOKETRYNTHQPEGC 506
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 431 LGBIRVNETNQCLDNNWARKENEKVG---IFNCHGMGNQVFSYTANKETIRD-----DLC 482
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 507 IAYEAGMDTLIMHLCETAPENO-----KFTLQEDGSLPHEBSKCKVQAARKESDSF 559
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 483 LDVSKLNGPVTMLKCHH-LKGNQLWEYDVPVKLTQ-----HVNNSQCLDKATDE--DSQ 533
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 560 VPLLROCTNSDHQKWFVK 577
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 534 VPSIRDCSGSRSQOVLRL 551
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 15
US-10-205-219-36
; Sequence 36, Application US/10205219
; Publication No. US20030138803A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
; APPLICANT: Brooksbank, Robert
; APPLICANT: Pinnock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WL-A-018200
; CURRENT APPLICATION NUMBER: US/10/205,219
; CURRENT FILING DATE: 2002-07-24

```


Result No.	Query	Score	Match	Length	DB	ID	Description
1	1194.5	38.2	623	2	T42245	probable polypepti	
2	1176	37.6	626	2	T42246	polypeptide N-acet	
3	1173	37.5	624	2	T42247	polypeptide N-acet	
4	1156.5	37.0	633	2	JC5247	polypeptide N-acet	
5	1101.5	35.3	559	2	JC4223	polypeptide N-acet	
6	1100.5	35.2	559	2	A45987	polypeptide N-acet	
7	1013.5	32.4	571	2	I37405	polypeptide N-acet	
8	1002	32.1	563	2	A88515	polypeptide N-acet	
9	1002	32.1	612	2	T42243	probable polypepti	
10	1001	32.0	589	2	T42244	probable polypepti	
11	999	32.0	617	2	T42249	polypeptide N-acet	
12	988.5	31.6	579	2	T31549	polypeptide N-acet	
13	973.5	31.2	618	2	T42248	polypeptide N-acet	
14	949	30.4	562	2	T42250	polypeptide N-acet	
15	839.5	26.9	601	2	T42251	polypeptide N-acet	
16	808.5	25.9	605	2	T27397	hypothetical prote	
17	789.5	25.3	684	2	T26930	hypothetical prote	
18	705	22.6	421	2	T42252	polypeptide N-acet	
19	299.5	9.6	276	2	T12552	hypothetical prote	
20	169.5	5.4	318	2	D87506	glycosyl transfera	
21	144.5	4.6	972	2	T09595	glucuronosyltransf	
22	142.5	4.6	297	2	D69327	glucuronoglycan bios	
23	141	4.5	313	2	A12404	hypothetical prote	
24	141	4.5	749	2	A95953	probable bifunctio	
25	138	4.4	312	2	S74669	hypothetical prote	
26	135	4.3	309	2	B87550	glycosyl transfera	
27	132.5	4.2	302	2	A84263	hypothetical prote	
28	128.5	4.1	328	2	B84263	succinoglycan bios	
29	124	4.0	308	2	H87306	glycosyl transfera	

QY	51	GPRTPRP-----GRREP-----VMRPPV-----PANALGARGEAV---RLQ	85
Db	66	GPPIEPBPVVENNKVEBEEQPGGNLAKPKFMDVNDPIVKKGDAQAQELGKAVVDDTK	125
QY	86	LQGBELRLQEBESVRLHQINLYLSDRISLHRRILPERNNPLCKEKKYDYNLPTSIIAPY	145
Db	126	LSTBEKAKYDKGMUNAFNAQYADMSIVHRTLPTNIDABCKTEKYN-ENLPTSIIICFH	184
QY	146	NEAWSTLLRTVYSVLETSPOILLEEVLDVYSDREHLKERLANELSGL-PKVRILIRANK	204
Db	195	NEANSVLLRTHVSVLERTPHLLLEEVLVDDFSDMDHTKRPLEEYMSQGGKVILRWK	244
QY	205	REGLVRLRLGASAAAGDVTLTFDHCCECHGWLPLELQRIHEBSAVVCPVIDIDWNT	264
Db	245	REGILRLRGAAVATGEVTLTLDSHCECMEGMWPELDRIRKDPRTTVVCPVIDIDNT	304
QY	265	FEYLGNSGE-PQJGGFDWRLVFTWHTVPRERIRMQSPVDVTRSPMTAGGLFAVSKYPE	323
Db	305	FEYHHSKAYFTSVGGFPDWGLQFNWHSIPERDKRNKTRTDPVRSPTMAGGLFSIDKYPE	364
QY	324	YLGSYDTGMEVWGGENLEFSFRIWQCGGVLETHPCSHVGHVFPKQAPY-----SRNKALAN	379
Db	365	KLGTIYDPGFDIWGGENLELSFKIWMCGGTLEIYPCSHVGHVFKRSPYKWRITGVNVLRN	424
QY	380	SVRAAEVMMDEFKELYHHRNPRARLEBFGDVTERKOLDKLOCKDFKWFLETYYPELVHP	439
Db	425	SIRLAEVMDDYKTYTYER-INNOJGDFGDISRKKLRDLGCKSPKMYLDNIYPELVFP	483

A;Reference number: JC4223; MUID:96115928; PMID:8690719
A;Accession: JC4223
A;Molecule type: mRNA
A;Residues: 1-559 <MEU>
A;Experimental source: salivary gland
A;Note: The authors translated the codon AAT for residue 264 as Asp
R;White, T.; Bennett, E.P.; Paul, E.; Takio, K.; Srensen, T.; Bonding, N.; Clausen, H.
J. Biol. Chem. 270, 24156-24165, 1995
A;Title: Purification and cDNA cloning of a human UDP-GalNAC:polypeptide N-Acetylgalactosyltransferase
A;Reference number: I37404; MUID:96025800; PMID:7592619
A;Accession: I37404
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-559 <RES>
A;Cross-references: EMBL:X85018; NID:G971458; PID:CAA59380.1; PID:G971459
C;Comment: This enzyme catalyzes the initial reaction in O-linked (mucin type) oligosaccharide biosynthesis.
C;Genetics:
A;Map position: 18
C;Superfamily: polypeptide N-acetylgalactosaminyltransferase
C;Keywords: chromosomal protein; glycoprotein; glycosyltransferase; hexosyltransferase; sialyltransferase
F;9-28/Domain: transmembrane #status predicted <TM>
F;29-559/Domain: endoplasmic reticulum lumenal #status predicted <LUM>
F;95,141,541,552/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;117,118,288/Binding site: carbohydrate (Thr) (covalent) #status predicted
F;119/Binding site: carbohydrate (Ser) (covalent) #status predicted

Query Match 35 34; Score 1101.5; DB 2; Length 559;
Best Local Similarity 43.04; Pred. No. 7.2e-77;
Matches 240; Conservative 86; Mismatches 171; Indels 61; Gaps 17;

QY 38 RAORGAGAGAAEPGPRTPRGREPPVPRPPVPANALGARGEAARLQGEELRQEE 97
DB 37 KKERGLPAGDV-----LEPVQK-----PHEGEGMGKPVVIPEDQZ--RWKEM 78
QY 98 VRLHQINILSDRTSLHRRLLPERWNPLCKEKYVDNLPRTSVIIAFYNEMSTLRTVY 157
DB 79 FKINQFNLMASEMIALNRSIPDVRLEGCKTKVP-DNLPTTSVVI VFHNEAWSTLLRTVH 137
QY 158 SVLETSPIILLEEVILVDDYSDEHKLKRLANELSGLP-KVPLIRANKREGLVARLLGA 216
DB 138 SVNRSRPHMIEEIVLVDASERDFLRKPLSEYVKKLKVPVHVIRMEQSRGLIRALKGA 197
QY 217 SAARGDVLTFDCHCECHGMLPLLQRIHEESAVVCPVIDVIDWNTFEYLGNSGEPQI 276
DB 198 AVSKGQVITFLDAHCETCTVGHLEPLLARIKHDRRTVVCPIIDVISDDTFEYMAGS--DMTY 256
QY 277 GGFDRRLVFTWHTVPERIRIMQSPDV-IRSPTMAGGLFAVSKKYFEYLGSDYTGMEVW 335
DB 257 GGFNWLKPRWYPVQREMRRKGRDRTLPRVPTTMAGGLFSIDRDYFQEGITGYDAGMDIW 316
QY 336 GGENLESPRIWQCGGVLETHPCSHVGHVFPQAPYS-----RNKALANSVRAAEVW 387
DB 317 GGENLEISFRWQCGGTLEIVTCSHVGHVFPKATYPTFGGTGQIINK---NNRRLAEVW 373
QY 388 MDEFKELYHYRNPRLRPPGDDVTERRKLRDLCKQDFKWFLETYVPELVPEDRPGFFG 447
DB 374 MDEFKQFFYIISPGVTVDYVDGDISRVCGLRHLCKLPFSWYLENIYPSQIIPH---YFS 430
QY 448 MLQNKGL-TDYCFDYNPDENQIVGHQVILYCHGMGNQGFETYSQKEIRYNTHOPEGC 506
DB 431 LGEIRNVETNQCLDNMARKENKVG---IFNCHGMGNQVFSYANKERTD---DLC 482
QY 507 IAYEAGMDTLIMLCETAPENQ-----KFLEQDGSFLFHEQSKKCVQAARKESDSF 559
DB 483 LDVSKLNGPVTMLKCHH-LKGNQLWEYDPVKLTQ-----HVNQNQCCLKATEE---DSQ 533
QY 560 VPLLRDCTNSDHQWFFK 577
DB 534 VPSIRDNGSRSSQWLLR 551

RESULT 6

Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biology
A;Reference number: A75000; MUID:99069613; PMID:981916
A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999, and
A;Accession: A88515
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-563 <STO>
A;Cross-references: GB:chr_III; PIDN:AAA28224.1; PID:g289776; GSPDB:GN00021
C;Genetics:
A;Map position: 3
C;Superfamily: polypeptide N-acetyl galactosaminyltransferase

```
Query Match      32.1%; Score 1002; DB 2; Length 563;
Best Local Similarity 41.0%; Pred. No. 3.4e-69;
Matches 227; Conservative 90; Mismatches 178; Indels 58; Gaps 18;

Qy 38 RAQAGAGAAERPPPTP-----RGRRE-PVMPRPVPPANAL----- 75
Db 4 RARR-----IEPLPAAQHSDPDAPHOPEKQKQVYPVDKETANQLKLMETQAFGP 57

Qy 76 GARGE-AVRLOQGEELRLOESVRLHQINILSDRLSHRRLPERNPLCKEKKYDYD- 133
Db 58 GYHGOGGTGVTPEDKKTIEKEFLENQFNVAASEMISVNRITLDPYSDACRTSGNNLKT 117

Qy 134 -NLPRTSVIIAFNEAWSTLLRTVYSVLETSPIILLEEVILVDDYSDREHLKERLANELS 192
Db 118 AGPKTSIIIVFNEAWTLLRTLHVSINRSPRLLEEIIIVDDKSDRYLVKPLDSYIK 177

Qy 193 GLP-KVRLIRANKREGVLRAALGASARGDVLTFDCHCEHGWLEPLQLRIHEESA 251
Db 178 MFPIPIHLVHLENRSGILRAALTGSEMAKGIILFLDAHVEVTDGWLPLEVSRVAEDRKR 237

Qy 252 VVCPVIDVDWNTFEYLGNSGEPOIGGDFMRLVFTWHTVPERERIRMQSPDVV-IRSPTM 310
Db 238 VVAPIIDVIDDITFEYV-TASSETWGGFNHNLFRWYAVKRELNRGRSDRSMPIQTPTI 296

Qy 311 AGGLFAVSKKYFEYLGSDYTGMEVWGGENLEFSPRIWQCGVLETHPCSHVGHVFPKQAP 370
Db 297 AGGLFAIDKQFFYDYGSDGQVWGGENLEISFRVWMCGLSLEIHFPCSRVGHVFRKQTP 356

Qy 371 YSRNKALA-----NSVRAAEVWMDPEKELYHYRNPRLARLEPFGDVTERRKOLRKLCKDF 425
Db 357 YTFPGGTAKVIHNAARTAEVWMDYKAFYKMWPAARNVAEAGDVSERKKLRETLOCKSF 416

Qy 426 KWLETVYPELVHPEDRPGFFGLMQLNGKLTIDYCFDYNPPDENQIVGHQVILYLCHGMGN 485
Db 417 KWLENLYPEALPADFRS-LGAIIVNR-FTEKCVDTNGKKDQAPGIAQ-----CHGAGGN 470

Qy 486 QFFEYTSQKEIRYNTHOPEGCIA-----VEAGMDTLIMHLCEETAPENQKFILOED---GS 538
Db 471 QANSLTGKGEIR-----SDDLCLSSGHVYQIGSB-LKLERC-SVSKINVGHVFFDDOAGT 524

Qy 539 LFHEQSKKCVQAA 551
Db 525 LLHKTKGKCVTGA 537
```

RESULT 9

T42243
probable polypeptide N-acetyl galactosaminyltransferase (EC 2.4.1.41) - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Nov-2000
C;Accession: T42243; S44913
R;Hagen, F.K.; Nehrke, K.
J. Biol. Chem. 273, 8268-8277, 1998
A;Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypept
A;Reference number: Z22126; MUID:98192620; PMID:9525933
A;Accession: T42243
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-612 <HAG>

A;Cross-references: EMBL:AF031833; NID:g3047186; PIDN:AAAL3669.1; PID:g3047187
R;Wilson, R.
submitted to the EMBL Data Library, May 1993
A;Description: Sequence of the *C. elegans* cosmid ZK688.
A;Reference number: S44913
A;Accession: S44913
A;Molecule type: DNA
A;Residues: 50-612 <WIL>
A;Cross-references: EMBL:LI6621; NID:g289775; PIDN:AAA28224.1; PID:g289776
C;Genetics:
A;Gene: gly-3
A;Introns: 101/1; 229/3; 520/3; 570/3
C;Superfamily: polypeptide N-acetyl galactosaminyltransferase
C;Keywords: glycosyltransferase; hexosyltransferase

```
Query Match      32.1%; Score 1002; DB 2; Length 612;
Best Local Similarity 41.0%; Pred. No. 3.8e-69;
Matches 227; Conservative 90; Mismatches 178; Indels 58; Gaps 18;

Qy 38 RAQAGAGAAERPPPTP-----RGRRE-PVMPRPVPPANAL----- 75
Db 53 RARR-----IEPLPAAQHSDPDAPHOPEKQKQVYPVDKETANQLKLMETQAFGP 106

Qy 76 GARGE-AVRLOQGEELRLOESVRLHQINILSDRLSHRRLPERNPLCKEKKYDYD- 133
Db 107 GYHGOGGTGVTPEDKKTIEKEFLENQFNVAASEMISVNRITLDPYSDACRTSGNNLKT 166

Qy 134 -NLPRTSVIIAFNEAWSTLLRTVYSVLETSPIILLEEVILVDDYSDREHLKERLANELS 192
Db 167 AGPKTSIIIVFNEAWTLLRTLHVSINRSPRLLEEIIIVDDKSDRYLVKPLDSYIK 226

Qy 193 GLP-KVRLIRANKREGVLRAALGASARGDVLTFDCHCEHGWLEPLQLRIHEESA 251
Db 227 MFPIPIHLVHLENRSGILRAALTGSEMAKGIILFLDAHVEVTDGWLPLEVSRVAEDRKR 286

Qy 252 VVCPVIDVDWNTFEYLGNSGEPOIGGDFMRLVFTWHTVPERERIRMQSPDVV-IRSPTM 310
Db 287 VVAPIIDVIDDITFEYV-TASSETWGGFNHNLFRWYAVKRELNRGRSDRSMPIQTPTI 345

Qy 311 AGGLFAVSKKYFEYLGSDYTGMEVWGGENLEFSPRIWQCGVLETHPCSHVGHVFPKQAP 370
Db 346 AGGLFAIDKQFFYDYGSDGQVWGGENLEISFRVWMCGLSLEIHFPCSRVGHVFRKQTP 405

Qy 371 YSRNKALA-----NSVRAAEVWMDPEKELYHYRNPRLARLEPFGDVTERRKOLRKLCKDF 425
Db 406 YTFPGGTAKVIHNAARTAEVWMDYKAFYKMWPAARNVAEAGDVSERKKLRETLOCKSF 465

Qy 426 KWLETVYPELVHPEDRPGFFGLMQLNGKLTIDYCFDYNPPDENQIVGHQVILYLCHGMGN 485
Db 466 KWLENLYPEALPADFRS-LGAIIVNR-FTEKCVDTNGKKDQAPGIAQ-----CHGAGGN 519

Qy 486 QFFEYTSQKEIRYNTHOPEGCIA-----VEAGMDTLIMHLCEETAPENQKFILOED---GS 538
Db 520 QANSLTGKGEIR-----SDDLCLSSGHVYQIGSB-LKLERC-SVSKINVGHVFFDDOAGT 573

Qy 539 LFHEQSKKCVQAA 551
Db 574 LLHKTKGKCVTGA 586
```

RESULT 10

T42244
probable polypeptide N-acetyl galactosaminyltransferase (EC 2.4.1.41) - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C;Accession: T42244
R;Hagen, F.K.; Nehrke, K.
J. Biol. Chem. 273, 8268-8277, 1998
A;Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypept
A;Reference number: Z22126; MUID:98192620; PMID:9525933
A;Accession: T42244
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA

A;Residues: 1-589 <HAG>
A;Cross-references: EMBL:AF031834; NID:g3047188; PIDN:AAC13670.1; PID:g3047189
C;Genetics:
A;Gene: gly-4
C;Superfamily: polypeptide N-acetylglalactosaminyltransferase
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 32.0%; Score 1001; DB 2; Length 589;
Best Local Similarity 38.5%; Pred. No. 4.3e-65;
Matches 232; Conservative 100; Mismatches 215; Indels 56; Gaps 18;

Qy 11 PRELRGREALVLLALLAGLSVLRQAQAGAGAAEPG--PPRTPR-----PGR 60
Db 3 PRMLK--MKTGTVLAVLWLEGLAFY--VQSTSSLRPPGRHPPPLQPLDLPQNPQ 58
Qy 61 REPVMPR---PPVPA-----NALGARGEA-----VRLQGLBEIRLQESVRLHQI 103
Db 59 NDEIRPKSAPPIPTINLAEDTTIHERTEKDVTKTFDEKFLNKGWQHOGEDKYKANSF 118
Qy 104 NIYLSDRISLHRLRPERWNPCKEKKYDYNLPRTSVIIAFYNEAWSTLRTVYVLETS 163
Db 119 NQASDALNPTRKIPRSREPOCRDVSQVGMQPTTIIITHNARSLLRTVPSVFNQS 178
Qy 164 PDILLEEVILVDDYSREHLKERLANELSGLPKRLIRANKREGVLRARLIGASAAAGDV 223
Db 179 PEELLEIVLVDNSQ----DVEIGKELAQIRITVLNNQREGVLRARLIGASAAAGDV 234
Qy 224 LTFDCHCEHEGWLEPLQRIHEESAVVCPVIDIDWNTFEYLGNSGPGPQIGFQWRL 283
Db 235 LTFDLSHIECNQKWLPLELARIENKAVAPIIDVINNFVNGASADLR--GFPDWTL 293
Qy 284 VFTWHTVPER--ERTMSPVDVIRSPMAGLFAVSKYFEYLGSDYTGMEVNGENLEF 342
Db 294 VFRFEMNEQURKERHAHTAPISPTMAGLFAISKEWFNELGTVDLDMEVNGENLEM 353
Qy 343 SFRVQCGVLETHPCSHVGHVFPKQAPY-----SRNKALANSVRAAEVWMDPEKELYH 397
Db 354 SFRVQCGVLETHPCSHVGHVFPKQAPY-----SRNKALANSVRAAEVWMDPEKELYH 413
Qy 398 RNPRLPEPGDVTERRKQKLRKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 457
Db 414 NVPSARFVNGDITDRLAIRDLQCKSPKYLENNVYPOLEIPRKTTPG--KSFQMK--IGNL 470
Qy 458 CFYDNPDPENQVGHVLYLCHGNGONQFYEYSQKIRYNTHQPEGCIAGEAGMD--T 515
Db 471 CLDSVARKESAPG----LFGCHGTGGNQEWVF--DQLTKTKFNALISQLCLDFSSNTENKT 525
Qy 516 LIMHLCETAPENQKFIQEDGSLFHEQSKCQVQAARKESSDSFVPLLRDCTNSDHQKWF 575
Db 526 VTMVKCENLRPDT--MVVEKNGWL--TQGGKLTWNGSGGDLIYGAHCELNGAQRWI 581
Qy 576 FKE 578
Db 582 FEK 584

RESULT 11
T42249
polypeptide N-acetylglalactosaminyltransferase (EC 2.4.1.41) 6b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C;Accession: T42249
J. Hagen, F.K.; Nehrke, K.
J. Biol. Chem. 273, 8268-8277, 1998
A;Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypept
A;Reference number: 222126; MUID:98192620; PMID:9525933
A;Accession: T42249
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-617 <HAG>
A;Cross-references: EMBL:AF031839; NID:g3047198; PIDN:AAC13675.1; PID:g3047199
C;Genetics:
A;Gene: gly-6

C;Superfamily: polypeptide N-acetylglalactosaminyltransferase
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 32.0%; Score 999; DB 2; Length 617;
Best Local Similarity 42.5%; Pred. No. 6.5e-69;
Matches 224; Conservative 80; Mismatches 149; Indels 74; Gaps 18;

Qy 71 PANALGARGEAVRLQGLBEIRLQESVRLHQIYLSDRISLHRLRPERWNPCKEKKY 130
Db 94 PHDDWEGGAGVS--HITPEQOKLADSTFAVNQFNPLVSDGISVRSLSPEIRKSCRNMVY 152
Qy 131 DYDNLPTSVIIAFYNEAWSTLRTVYVLETSVSPDILLBEVILVDDYSREHLK--ERLAN 189
Db 153 P-DNLPTSVIIYHNEAVSTLRTVMSVIDRSPKELLKEIILVDDFSDFREFLRYPTLDT 211
Qy 190 ELSGLP--KVLIRANKREGVLRARLIGASAAAGDVLTFLDCHCEHEGWLEPLLRHES 248
Db 212 TLKPLPTDIKIIRSKERVGLIRAMWGAQAQGDVLTFLDCHCECTKGWLEPLLRHES 271
Qy 249 ESAVVCVIDIDWNTFEYLGNSG--EPQIGGFQWRLVFTWHTVP--ERERIRMQSPVDVIR 306
Db 272 RKAVCPVIDIDWNTFEYLGNSG--EPQIGGFQWRLVFTWHTVP--ERERIRMQSPVDVIR 329
Qy 307 SPTMAGLFAVSKYFEYLGSDYTGMEVNGENLEFSFRIWQCGVLETHPCSHVGHVFP 366
Db 330 SPTMAGLFSINRYFEELGEYDPGMDIINGENLEMSFRIWQCGVLETHPCSHVGHVFP 389
Qy 367 KQAPY-----SRNKAL--ANSVRAAEVWMDPEKELYHNPRA--RLSPFGDVTERRKQK 419
Db 390 KSPHDPFGKSGKVLNLLRVAEVMWMDKHGYFYKIAFOAHMRSSIDVSRVEURKK 449
Qy 420 LQCKDFKWLFTVYPE--LHVPEDRPGFQMLQNGLTDYCFDYNP----PDENQIVGHQ 473
Db 450 LNCCKSFNLYQNVFQHFLEPTLDR---FORMSN---SNYCTAFPGDTGPKHRLUGSP 503
Qy 474 VILYCHGNGONQFYEYSQKIRYNTHQPEGCIAGEAGMDTILMHLCEETAPENQKFI 533
Db 504 CTM----GPDWLQWLVTGDRIRTDHLL---CLSVV-----LKECAGFDTEYWDKPKI 533
Qy 534 QEDGSLFHEQSKCQVQAARKESSDSFVPLLRDCTNSDHQKWFKERM 580
Db 534 ----QLLHTTSDWKIQ-----LKECAGFDTEYWDKPKI 563

RESULT 12

T31549
polypeptide N-acetylglalactosaminyltransferase (EC 2.4.1.41) 9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 01-Dec-2000
C;Accession: T31549; T42253
R. Matthews, L.
submitted to the EMBL Data Library, October 1999
A;Reference number: 221043
A;Accession: T31549
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-579 <WIL>
A;Cross-references: EMBL:AL117202; PIDN:CAB57897.1; CESP:Y47D3A.23
A;Experimental source: Clone Y47D3A
J. Hagen, F.K.; Nehrke, K.
J. Biol. Chem. 273, 8268-8277, 1998
A;Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypept
A;Reference number: 222126; MUID:98192620; PMID:9525933
A;Accession: T42253
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-202,'D',204-579 <HAG>
A;Cross-references: EMBL:AF031843; NID:g3047206; PIDN:AAC13679.1; PID:g3047207
C;Genetics:
A;Gene: CESP:Y47D3A.23; gly-9
A;Introns: 45/1; 179/1; 276/1; 343/3; 373/1; 453/3; 491/3; 533/3
C;Superfamily: polypeptide N-acetylglalactosaminyltransferase
C;Keywords: glycosyltransferase; hexosyltransferase

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 19:39:32 ; Search time 17 Seconds
(without alignments)
1607.207 Million cell updates/sec

Title: US-10-074-527-2

Perfect score: 3124

Sequence: 1 MWGRTARRRCPRELRRGREA.....LLRDCNTNSHQKWFKKERML 581

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1109	35.5	559	1	PAGT_MOUSE
2	1101.5	35.3	559	1	PAGT_HUMAN
3	1100.5	35.2	559	1	PAGT_BOVIN
4	1097	35.1	559	1	PAGT_PIG
5	1097	35.1	559	1	PAGT_RAT
6	1002	32.1	612	1	PAG3_CAEAL
7	121.5	3.9	755	1	BPHY_DEIRA
8	119.5	3.8	348	1	EXOO_RHME
9	118.5	3.8	294	1	Y696_HAEIN
10	116	3.7	322	1	Y586_ANASP
11	112.5	3.6	1618	1	NEST_HUMAN
12	112.5	3.6	3038	1	TRIO_HUMAN
13	110	3.5	428	1	GGPP_NEUCR
14	110	3.5	477	1	YXNA_STRLI
15	109.5	3.5	1275	1	RFBC_MYXXA
16	108.5	3.5	909	1	Y4GI_RHISN
17	108	3.5	513	1	YTH1_RHOER
18	105.5	3.4	301	1	AMSB_ERWAM
19	103	3.3	445	1	CPWD_HUMAN
20	102.5	3.3	1001	1	TP3A_HUMAN
21	102	3.3	267	1	DPM1_YEAST
22	102	3.3	1527	1	CAIH_MOUSE
23	99.5	3.2	1919	1	HAPP_RAT
24	99	3.2	250	1	Y868_HAEIN
25	99	3.2	589	1	PHBC_ALCEU
26	98.5	3.2	344	1	YIBD_ECOLI
27	98	3.1	431	1	NOCT_HUMAN
28	96	3.1	419	1	HASA_STRPY
29	96	3.1	488	1	BTB1_MOUSE
30	95.5	3.1	900	1	GGAB_BACSU
31	95	3.0	500	1	U713_HUMAN
32	94.5	3.0	1118	1	CARB_YEAST
33	94	3.0	479	1	VGLC_PRVIF

34	94	3.0	528	1	SYK_TREPA
35	94	3.0	1230	1	RPOM_HUMAN
36	93.5	3.0	623	1	Y711_HUMAN
37	93.5	3.0	2470	1	TOR1_YEAST
38	93	3.0	471	1	RTN2_MOUSE
39	93	3.0	478	1	XJF4_YEAST
40	93	3.0	895	1	MCN2_HUMAN
41	93	3.0	904	1	MCN2_MOUSE
42	93	3.0	1663	1	HAPI_HUMAN
43	92.5	3.0	609	1	UGST_ORYGL
44	92.5	3.0	609	1	UGST_ORYSA
45	92.5	3.0	1455	1	DP2L_PYPAB

ALIGNMENTS

RESULT 1

PAGT_MOUSE	ID	PAGT_MOUSE	STANDARD;	PRT;	559 AA.
AC	O08912;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Polypeptide N-acetylglucosaminyltransferase				
DE	UDP acetylglucosaminyltransferase (EC 2.4.1.41) (Protein-				
DE	acetylglucosaminyltransferase) (UDP-GalNac-T1) (ppGantase-T1).				
GN	GALNT1.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.				
RX	MEDLINE=97298094; PubMed=9153242;				
RA	Hagen F.K., Ten Hagen K.G., Beres T.M., Balys M.M.,				
RA	Vanwuyckhuysen B.C., Tabak L.A.;				
RT	"CDNA cloning and expression of a novel UDP-N-acetyl-D-				
RT	galactosamine:polypeptide N-Acetylglucosaminyltransferase.";				
RL	J. Biol. Chem. 272:13843-13848(1997).				
CC	-I- FUNCTION. THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED				
CC	OLIGOSACCHARIDE BIOSYNTHESIS, THE TRANSFER OF AN N-ACETYL-D-				
CC	GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE				
CC	PROTEIN RECEPTOR. PAGT AND PAG4 HAVE DIFFERENT SUBSTRATE				
CC	SPECIFICITY.				
CC	-I- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =				
CC	UDP + N-acetyl-D-galactosaminyl-polypeptide.				
CC	-I- COFACTOR: MANGANESE AND CALCIUM (BY SIMILARITY).				
CC	-I- PATHWAY: Glycosylation.				
CC	-I- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.				
CC	-I- TISSUE SPECIFICITY: HIGHER EXPRESSION IN KIDNEY, HEART, SMALL				
CC	INTESTINE AND CERVIX AND TO A LESSER EXTENT IN ALL THE OTHER				
CC	TISSUES TESTED.				
CC	-I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.				
CC	-I- SIMILARITY: Contains 1 ricin B-type lectin domain.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; U73820; AAB58477.1; --				
DR	MGD; MGI:894693; Galnt1.				
DR	GO; GO:0004653; F:polypeptide N-acetylglucosaminyltransferase. . . ; IDA.				
DR	GO; GO:0006493; P:O-linked glycosylation; IDA.				
DR	InterPro; IPR001173; Glyco trans 2.				
DR	InterPro; IPR000772; Ricin_B lectin.				
DR	Pfam; PF00555; Glycos_transf_2; 1.				
DR	Pfam; PF00652; Ricin B lectin; 3.				
DR	SMART; SM00458; RICIN; 1.				


```
Db 79 FKIQNLMASEMIALNRSPLDVRLEGCKTKVYP-DNLPTTSVIVFHEAWSTLLRTVH 137
Qy 158 SVLETSDDLLEEVILVDDYSDRHLKRLERLANELSLGLP-KVRLIRANKREGLVRLRLLGA 216
Db 138 SVINRSPRMLLEETVLDVDDASERDFLRPLESVYKLLKVPVHVIRMEQSRGLIRARLKGA 197
Qy 217 SAARGDVLTFDCHCECHGWLPLELQRIHEESAVVCPVDIVDMNTFFVLGNSGSPQI 276
Db 198 AVSGQVITFDLHAECCTVGLWLEPLARIKHDKRTVVCPIIDVSDDTFEYMAGS-DMTY 256
Qy 277 GGFDRVLVFTVHTVPERIRMQSPVDV-IRSPTMAGGLFAVSKKYPFYLGSYDTGMEVW 335
Db 257 GGFNKLNFWRVYVPQREMDRRKGRDRTLVRTPTMAGGLFSIDRDYFQIGTYDAGMDIW 316
Qy 336 GGENLEFSRIWQCGGVLETHPCSHVGHVFPKQAPYS-----RNKALANSVRAAEVW 387
Db 317 GGENLEFSRIWQCGGTLEIVTCSHVGHVFRKATPYTPPGTGTQIINK---NNRRLAEVW 373
Qy 388 MDEKELYHNRPRARLEPFGDVTERRKQDKLOCKDKWLETVVPELHVPEDRPGFG 447
Db 374 MDEFKNFYIISPGTVKDYGDISRLGLRHLKQCRPFSWYLENIYDPSQIPRH---YFS 430
Qy 448 MLQNKGL-TDYCFDYNPPDENQIVGHQVILYLCHGMGQNPFFYTSQKEIRYNTHOPEGC 506
Db 431 LGEIRNVETNQLDNMARKENKVG----IFNCHGMGNGQVFSYANKIERTD---DLC 482
Qy 507 IAVBAGMDTLIMHLCETAPENQ-----KFIQEDGSLFHEQSKKCVQAKRESDSF 559
Db 483 LDVSKLNGPVTMLKCHH-LKGNQLWEYDPVKLTQ-----HVNNSQCLDKATDE--DSQ 533
Qy 560 VPLARDCTNSDHQKWF 577
Db 534 VPSIRDCSGRSQQWLLR 551
```

RESULT 4

```
PAGT_PIG
ID PAGT_PIG STANDARD; PRT; 559 AA.
AC Q29121;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) (Protein-
DE UDP acetylglucosaminyltransferase) (UDP-GalNac:polypeptide, N-
DE acetylglucosaminyltransferase) (GalNac-TL).
GN GALNT1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxId=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=96318021; PubMed=8748160;
RA Yoshida A., Hara T., Ikenaga H., Takeuchi M.;
RL Glycoconj. J. 12:824-828(1995).
CC -!- FUNCTION: THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED
CC OLIGOSACCHARIDE BIOSYNTHESIS, THE TRANSFER OF AN N-ACETYL-D-
CC GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE
CC PROTEIN RECEPTOR.
CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =
CC UDP + N-acetyl-D-galactosaminyl-polypeptide.
CC -!- COFACTOR: MANGANESE AND CALCIUM (BY SIMILARITY).
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC ENBL; D85389; BAA12800.1; --
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00535; Glycos_transf_2; 1.
DR Pfam; PF00652; Ricin_B_lectin; 3.
DR SMART; SM00458; RICIN_1.
DR PROSITE; PS0231; RICIN_B_LECTIN; 1.
DR Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW Golgi stack; Glycoprotein; Manganese; Calcium; Lectin.
FT REMOVED IN SOLUBLE POLYPEPTIDE N-
FT PROPEP 1 40
FT ACETYLGLACTOSAMINYLTRANSFERASE (BY
FT SIMILARITY).
FT CHAIN 41 559
FT SOLUBLE POLYPEPTIDE N-
FT ACETYLGLACTOSAMINYLTRANSFERASE (BY
FT SIMILARITY).
FT DOMAIN 1 8
FT CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 9 28
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 29 559
FT LUMENAL, CATALYTIC (POTENTIAL).
FT DOMAIN 439 559
FT RICIN B-TYPE LECTIN.
FT CARBOHYD 95 95
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 117 117
FT O-LINKED (POTENTIAL).
FT CARBOHYD 118 118
FT O-LINKED (POTENTIAL).
FT CARBOHYD 119 119
FT O-LINKED (POTENTIAL).
FT CARBOHYD 141 141
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 288 288
FT O-LINKED (POTENTIAL).
FT CARBOHYD 552 552
FT N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 559 AA; 64118 MW; 68C2D261A51684C6 CRC64;
```

Query Match 35.1%; Score 1097; DB 1; Length 559;

Best Local Similarity 41.5%; Pred. No. 4.3e-77;

Matches 243; Conservative 84; Mismatches 175; Indels 84; Gaps 17;

```
Qy 9 RCPRELRGREALVLLALLALAGLSVLRQAGAGAAEPGPPTPRGRPEVNNRPP 68
Db 33 KCKEKGRLPA-----GDVLEPVQKPHGPGEMGKP-----VVIK- 69
Qy 69 PVPANALGARGEAVRLQOGEELRLOEESVRLHQINILYSDRLSLRRLPERWNPLCKEK 128
Db 70 -----EDQDKVMEKFNQFNLMASEMIALNRSPLDVRLEGCKTK 109
Qy 129 KYDYDNLPTTSVIAFYNEAWSTLIRTVYSLVSTSDILLEEVLVDYSDRHLKERLA 188
Db 110 VYP-DNLPTTSVIVFHEAWSTLLRTVHSVINRSPRMLLEEVILVDDASERDFLRKPLE 168
Qy 189 NELSGLP-KVRLIRANKREGLVRLRLLGASAAAGDVLTFDCHCECHGWLPLELQRIHE 247
Db 169 SYVKLLKVPVHVIRMEQSRGLIRARLKGAASVQVITFDLHAECCTVGLWLEPLARLKH 228
Qy 248 ESASAVCPVIDIVDMNTFFVLGNSGSPQIGFDRVLVFTVHTVPERIRMQSPVDV-IR 306
Db 229 DRKTVVCPVIDIVSDDTFEYMAGS-DMTYGGFNWKLNFWRVYVPQREMDRRKGRDRTLVR 287
Qy 307 SPTMAGGLFAVSKKYPFYLGSYDTGMEVWGENLEFSRIWQCGGVLETHPCSHVGHVFP 366
Db 288 TPTMAGGLFSIDRDYFQIGTYDAGMDIWGENLEISFRIWQCGGTLEIVTCSHVGHVFR 347
Qy 367 KQAPYS-----RNKALANSVRAAEVWMDPEKELYHNRPRARLEPFGDVTERRKQDK 418
Db 348 KATPYTPPGTGTQIINK---NNRRLAEVWMDPEKELYHNRPRARLEPFGDVTERRKQDK 404
Qy 419 KLQCKDFKWFLETVVPELHVPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYL 478
Db 405 KLQCRPFSWYLENIYDPSQIPRHYS--LGEIRNVETNQLDNMARKENKVG----IFN 458
Qy 479 CHGMGNGQVFSYANKIERTD-----DLCVDSKLNGLNGPVTMLKCHH-LKGNQLWEYDPVKL 531
Db 459 CHGMGNGQVFSYANKIERTD-----DLCVDSKLNGLNGPVTMLKCHH-LKGNQLWEYDPVKL 513
```

FT	CARBOHYD	141	141	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	288	288	O-LINKED (POTENTIAL).
FT	CARBOHYD	552	552	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	559 AA;	64229 MW;	5E36A95D9422C853 CRC64;

Query Match		35.1%;	Score 1097;	DB 1;	Length 559;	
Best Local Similarity		41.1%;	Pred. No. 4.3e-77;			
Matches 241;		Conservative	87;	Mismatches 173;	Indels 86;	Gaps 17;

Qy	9	RCPRELRRGREALVLLALLALAGLSVLRRAORAGAGAAEPGPRTPRGCRREPVMRPP	68
Db	33	KCEBKXERGLPA-----GDVLELVQKPEHGGEMKCP-----VVLPK-	69
Qy	69	PVPANALGARGEAVRLQOGEELRLOEESVRLHQINIIYLSDRISLHRRLLPERWNPLCKEK	128
Db	70	-----EDQEKWKEFKINQFNLMASEMIAFNRLSDVRLEGGCKTK	109
Qy	129	KYDYDNLPRTSVIAAFNEAWSLTLLRTVYSVLETSPOILLLEVLVDVDSYDREHLKERLA	188
Db	110	VYP-DSLPTTSVVIVFHNEAWSLTLLRTVRSVINRSPRHIEELVLVDVASERDFLKRPLE	168
Qy	189	NELSGLP-KVRLIRANKREGLVRARLLGASAARGDVLTFLDCHCECHGWLEPLLORIE	247
Db	169	SYVKLLKVPVHVIRMEQSRGLIRARKGAASVSGQVITFLDAHCECTVGWLEPLLARIKH	228
Qy	248	EEASVVCPIVDIWNTPFYLGNSGEPQIGGFQWRLVFTWHTVPERIRMQSPDVV-IR	306
Db	229	DRTVTVCPIIDVLSDDTFEYMAGS-DMTYGGFGFNKLFNRYVPVQREMDRRKDRDRLPVR	287
Qy	307	SPTMAGGLFAVSKKPYEYLGSDYDTGMVWGGENLEFSFRIWQCGGVLETHPCSHVGHVFP	366
Db	288	TPTMAGGLFSIDRDYFQEIGTYDAGWDIWGGENLEISFRIWQCGGTLIEIVTCSHVGHVFR	347
Qy	367	KQAPYS-----RNKALANSVRAAEVWMDPEFKELYHYHRNPRARLEPFQGDVTERKQLRD	418
Db	348	KATPYTFPGGTQGIINK--NNRRLAEVWMDPEFKNFYIISPGVTKVDYDGISRSGVLRH	404
Qy	419	KLOCKDPKFWLETPYELHVPEDRPGFFGMLQNKGL-TDYCFDYNPDENQIVGHQVILY	477
Db	405	KLOCKPFSWYLENIYDPDSQIPRH---YFSLGEIRNVETNOCLDNMARKENKRVG----	457
Qy	478	LCHGMQGNQFFEYTSOKERYNTHQPEGGIAVBAGMDTLIMHLCEBTAPENQ-----K	530
Db	458	NCHGMGNGVFSYTANKEIRTD---DLCLDVSUKLNGPVMTLCKHH-LKGNQLWEYDPVK	512
Qy	531	FILQEDGSLFHEQSKCKVQAARKESDSDSFVPLLRDCTNSDGHQWFKF	577
Db	513	LTLLQ-----HVNNSOCLDKATEE--DSQVPSTIRDTGSRSSQWLLR	551

```

RESULT 6
PAGE3_CAEEL
ID ID_PAG3_CABEL STANDARD; PRT; 612 AA.
AC P34678; Q9U003;
DT 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Polypeptide N-acetyl-galactosaminyltransferase 3 (EC 2.4.1.41)
DE (Protein-UDP acetyl-galactosaminyltransferase) (UDP-GalNAC:polypeptide,
DE N-acetyl-galactosaminyltransferase) (GalNAC-T1) (ppcANTase 3).
GN Glnf-3 OR ZK688-8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Bristol N2.
RC MEDLINE=98192620; PubMed=9525933;
RX Hagen F.K., Nehrke K.;
RA "cDNA cloning and expression of a family of UDP-N-acetyl-D-
RT galactosamine:polypeptide N-acetyl-galactosaminyltransferase sequence
RT homologs from Caenorhabditis elegans."

```

```

RL J. Biol. Chem. 273:8268-8277(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Suleston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL Nature 368:32-38(1994).
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED
CC OLIGOSACCHARIDE BIOSYNTHESIS, THE TRANSFER OF AN N-ACETYL-D-
CC GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE
CC PROTEIN RECEPTOR.
CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =
CC UDP + N-acetyl-D-galactosaminyl-polypeptide.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF031833; AAC13669.1; --
DR EMBL; U16621; AAR28224.3; --
DR PIR; T42243; T42243.
DR WormPep; ZK688.8; CE29649.
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00535; Glycos_transf_2; 1.
DR Pfam; PF00652; Ricin_B_lectin; 2.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
DR KX Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;
DR Glycoprotein; Lectin.
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 35 612 LUMENAL (POTENTIAL).
FT DOMAIN 478 612 RICIN B-TYPE LECTIN.
FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 612 AA; 68911 MW; 3031C2FE933F9858 CRC64;

Query Match 32.1%; Score 1002; DB 1; Length 612;
Best Local Similarity 41.0%; Pred. No. 1.1e-69;
Matches 227; Conservative 90; Mismatches 178; Indels 58; Gaps 18;

QY 38 RAQCAGAGAGAPGPRTP-----RGRRE-PVMPRPVPANAL-----75
DB 53 RARR-----TEPLPPAQHSDSDPAHPQPEKQKQVVDVDTANQLRKLMTQARGP 106
QY 76 GARGE-AVRLQGBELRLQESVRLHQINYLSDRISLHRLPWRMPLCKEKKYDYD- 133

```

```

107 GYHOGGTGVTVPEDKKTIKRFLNQFNVAASEMISVNRITLDPYRSACRTSGNNLKT 166
134 -NLPRTSVIAFYNEAWSTLLRTVYSVLETSPIILLEEVILVDDYSDREHLKERLANELS 192
167 AGPKTSIIIVFNEAWTTLRLTHSVINRSPHLLLEIILVDDKSDRDYLVKPLDSYIK 226
193 GLP-KVRLIRANKREGVLRLGSAARGDVLTLFDCHCEHEGWLPLELLQRIHEESA 251
227 MFPPIHLVHLENRSGILRLARLTGSEWAKGKILFLDAHVETDGLWLEPLVSRVAEDKR 286
252 VVCPVIDVDWNTFEYLGNSGEPOIGGFWRLLVFTWHTVPERERIRMQSPVDV-IRSP 310
287 VVAPIIDVIDSDTTFEYV-TASETTWGFENHLLNFRYAVPKRELNRGRSDRSMPIQPTI 345
311 AGGLEAVSKYFEYLGSDYTGMEVWGENLEFRIWOCGVLETHPCSHVGHVFPKQAP 370
346 AGGLEAIDKQFFDYDYGDMQVWGENLEIFRVMCGSLIHPCSRGVHVFRTQTP 405
371 YSRNKALA-----NSVRAAEVWMDPEFKELYHYHNRARLEPFGDVTERKOLRDKLCKDF 425
406 YTFPGGTAKVIHHAARTAEVWMDYKAFYKVPAPARNVEAGDVSEKRLRETLOCKSF 465
426 KWLETVYVPELVHPEDRPGFGLQNLQGLTDYCFDYNPPDENQIVGHQVILYLCHGMQON 485
466 KYLENIYPEAPLPADFRS-LGAIVNR-FTEKCVDTNGKXGQAPGIAQ-----CHGAGGN 519
486 QFPEYTSOKEIRYNTHOPEGCIA-----VEAGMDTLIMHLCETAPENKFILOED---GS 538
520 QAWSLTGKGEIR-----SDDLCLSGHYVQIGSE-LKLERC-SVSKINVKHVFVDDOAGT 573
539 LFHEQSKKCVQAA 551
574 LLHKTKGKCVTGA 586

RESULT 7
BPHY DEIRA
ID_BPHY DEIRA STANDARD; PRT; 755 AA.
AC Q9RZ4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bacteriophytochrome (EC 2.7.3.-) (Phytochrome-like protein).
GN BHPH OR DRA0050.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567286;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
RN [2]
RP CHARACTERIZATION, AND MUTAGENESIS OF MET-259; CYS-289 AND HIS-260.
RX MEDLINE=20085458; PubMed=10617469;
RA Davis S.J., Venter A.V., Vierstra R.D.;
RT "Bacteriophytochromes: phytochrome-like photoreceptors from
RT nonphotosynthetic eubacteria."
RL Science 286:2517-2520(1999).
CC -!- FUNCTION: PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT ARE
CC REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE R FORM THAT ABSORBS
CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE FR FORM THAT
CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. HAS ALSO A SLIGHT BLUE

```


Db 157 CAPLPAP-----PPAPAPPEV-----EELA-----178
Qy 111 ISLHRLPERWNPCKEKKYDNLPTSIIIAFYNEAWSTLLRTVSVLETSDDLLEE 170
Db 179 -----RRLGAWRG-----AVRGQE-----RVAH--METSLD-----204
Qy 171 VILVDDYSDBHLKERLANELSGLPKVLIRANKREGVLVARLLGASAGRDVLTFLDCH 230
Db 205 -----QTRERLARAVQAREVRL-----ELQLOAERGGLLERRAAL 241
Qy 231 CEEHEG-WLEPL-----LQRIHEESAVCFVIDVIDMNTFEYLGNSGEPIGGFD 280
Db 242 EORLEGRWOERLATERKPOLAVEALEQEKQGLQSOIAQVLE-----GRQOLAHLK 291
Qy 281 WRL---VPTWHTVERERIRQSPVDVIRPTMAGGLFAVSKKYFEVLGSLVDTGMEVWGG 337
Db 292 MSLSLEVATYRTLLEAENSRLQTP-----GGGSKTSLSPD-----327
Qy 338 ENLEFSFRIMQCGGVLETHPCSHGVHPFKOAPYSRNKALANSVRAAEVWMDPEKELYH 397
Db 328 PKLELQEP-----RTPEGRRLGSLPLVLSPTSLSPLEATLETVPAPFLKNQEFLOA 379
Qy 398 RNPRLARLEPPGQVTE-----RKQLRDLQCKDFKWFLETV 432
Db 380 RTPTLASTPIPTPQAPSPAVDAEIRAQDAPLSLLQTQGRKQAPELRAEARVAIPASV 439
Qy 433 YPELHVBPEDRGFGMLQKGLTDYCFDYN-----PPDENQI-----VCHQVILYLCH 480
Db 440 LPG---PBE-FG---GORQEAATGQSPEDHASLAPPLSPDHSLSRAKDGSGSVFISICR 493
Qy 481 GMGNQRFEXYSQKEIRYNTHQPSGCTAVEAGMTLIMHLCETAPENQKFIQEDGSLF 540
Db 494 GEGSGQIW-----GLVEKETAIGKVVSLSQEQIWE-----BED-----527
Qy 541 HEQSKVCQAARKSSDSFVPLLRDCTNS 569
Db 528 -----LNRKEIQDSQVLEKETLKS 547

RESULT 12
TRIO HUMAN
ID TRIO HUMAN STANDARD; PRT; 3038 AA.
AC 075962; Q13458;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Triple functional domain protein (PTRPF interacting protein).
GN TRIO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Fibroblast;
RX MEDLINE=96224308; PubMed=8643598;
RA Debant A., Serria-Pages C., Seipel K., O'Brien S., Tang M., Park S.-H.,
RA Streuli M.;
RT "The multidomain protein Trio binds the LAR transmembrane tyrosine
RT phosphatase, contains a protein kinase domain, and has separate rac-
RT specific and rho-specific guanine nucleotide exchange factor
RT domains".
RL Proc. Natl. Acad. Sci. U.S.A. 93:5466-5471 (1996).
RN [2]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RN Streuli M.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RN STRUCTURE BY NMR OF 1227-1407, AND MUTAGENESIS.
RX MEDLINE=99005194; PubMed=9790533;
RA Liu X., Wang H., Eberstadt M., Schnuchel A., Olejniczak E.T.,
RA Meadows R.P., Schkeryantz J.M., Janowick D.A., Harlan J.E.,

RA Harris E.A.S., Staunton D.E., Fesik S.W.;
RT "NMR structure and mutagenesis of the N-terminal Dbl homology domain
RL of the nucleotide exchange factor Trio.";
RN Cell 95:269-277 (1998).
RP [4]
RP CHARACTERIZATION.
RX MEDLINE=99276567; PubMed=10341202;
RA Seipel K., Medley Q.G., Kederasha N.L., Zhang X.A., O'Brien S.P.,
RA Serria-Pages C., Hemler M.E., Streuli M.;
RT "Trio amino-terminal guanine nucleotide exchange factor domain
RT expression promotes actin cytoskeleton reorganization, cell migration
RL and anchorage-independent cell growth.";
CC J. Cell Sci. 112:1825-1834 (1999).
CC -I- FUNCTION: PROMOTES THE EXCHANGE OF GDP BY GTP. TOGETHER WITH
CC LEUCOCYTE ANTIGEN-RELATED (LAR) PROTEIN, IT COULD PLAY A ROLE IN
CC COORDINATING CELL-MATRIX AND CYTOSKELETAL REARRANGEMENTS NECESSARY
CC FOR CELL MIGRATION AND CELL GROWTH.
CC -I- SUBUNIT: INTERACTS TO FORM A COMPLEX WITH LEUCOCYTE ANTIGEN
CC RELATED PROTEIN.
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=O75962-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O75962-2; Sequence=VSP_004467, VSP_004468;
CC -I- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART, SKELETAL MUSCLE,
CC BRAIN, PANCREAS, PLACENTA, LIVER, KIDNEY AND LUNG.
CC -I- DOMAIN: THE N-TERMINAL DBL/GEF DOMAIN SPECIFICALLY CATALYZES
CC NUCLEOTIDE EXCHANGE FOR RAC1, LEADING TO THE ACTIVATION OF JUN
CC KINASE AND THE PRODUCTION OF MEMBRANE RUFFLES. THE SECOND DBL/GEF
CC DOMAIN IS AN EXCHANGE FACTOR FOR RHOA AND INDUCES THE FORMATION OF
CC STRESS FIBERS.
CC -I- PTM: SERINE PHOSPHORYLATED.
CC -I- SIMILARITY: Contains 1 CRAL-TRIO domain.
CC -I- SIMILARITY: Contains 2 DBL-homology (DH) domains.
CC -I- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC -I- SIMILARITY: Contains 2 PH domains.
CC -I- SIMILARITY: Contains 1 SH3 domain.
CC -I- SIMILARITY: Contains 4 spectrin repeats.
CC -I- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SER/THR
CC FAMILY OF PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF0911395; AAC43042.1; -.
CC EMBL; U42390; AAC34245.1; -.
CC HSP; G63450; IAO6.
CC Genew; HGNC:12303; TRIO.
CC MIM; 601893; -.
CC GO; GO:0005085; F:guanyl-nucleotide exchange factor activity; TAS.
CC GO; GO:0004674; P:protein serine/threonine kinase activity; TAS.
CC GO; GO:0007185; P:transmembrane receptor protein tyrosine pho. . . ; TAS.
CC InterPro; IPR001251; CRAL-TRIO.
CC InterPro; IPR001331; GDS_CDC24.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003598; Ig_c2.
CC InterPro; IPR003006; Ig_YHC.
CC InterPro; IPR001845; PH.
CC InterPro; IPR000719; Prot kinase.
CC InterPro; IPR000219; RhoGEF.
CC InterPro; IPR002290; Ser thr_kinase.
CC InterPro; IPR001452; SH3_
CC InterPro; IPR002017; Spectrin.
CC Pfam; PF00047; Ig; 1.
CC Pfam; PF00169; PH; 2.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00621; RhoGEF; 2.

Query Match	3.6%	Score 112.5	DB 1	Length 3036
Best Local Similarity	19.2%	Pred. No. 5.7		
Matches 127	Conservative 96	Mismatches 211	Indels 229	Gaps 35
Qy	80	EAVRLIQGEE-----LRLOESVRLHQINIYLSDRI--SLHRRLLPERWNPLCK	126	
Db	883	EAEQLQREHQFOHAIEKTHQSALQVQQAEMLQANHYYDMIRDCAEKVASHWQQML	942	
Qy	127	EKKYVDNLPRTSVIIAFYNEAWSTLLRTVYSVLETSPDILLEEVILVDY-----	177	
Db	943	KME---DRLKLVASVAFKTS-----EQVCSVLES-----LQEKYREDWCAGDKLGP	990	
Qy	178	-SDRELKERLANELSGLPKVLRIANKREGLVRARLLGASAAAGDVTLFLD-----	228	
Db	991	NSETDHTVPMISGHL-----EQKEAFKKA-----CTLARNADVFLKYLHRSVN	1035	
Qy	229	-----CHCECHGWLPELILQRTHESESATV-----CPVIDVDWN-----TFEYL	268	
Db	1036	MPGVWTHIKAPEQOVKNILNELFORENRVLYHWTKRRRLDCCQVVFERSAKALEWI	1095	
Qy	269	GNSGEQIQGFQWRLVFTWHT-----VPERERIRM---QSPVDVIRSPTMAGG	313	
Db	1096	HDNGE-----FYLSTHTSGSSIQHTQELLKEHEFOITAKQTKERVKLLQLADG	1146	
Qy	314	-----LFAVSKKYFEY---LGSYDTGMEVWG--ENLEFSFRIMQCGVYL	353	
Db	1147	FCEKGHAHAAEIKKCVTAVDKRYRDFSLRMEKYRTSLEKALGISDSNKSLSQ---	1202	
Qy	354	ETHPCSHVG-HVFPKQAPYSRNKALANSVRAAEVWMDPEKELYHYHNPRLARLEP	412	
Db	1203	DIIPASIPGSEVKRLDAAHELNEERKKSARRKEFIMAEILQ-----TE	1245	
Qy	413	RKQRLDKLOCKD-FKWFLETVVPELVHVPDRPG-----PFGMLQN-----KGL	454	
Db	1246	KAVRDLRECDMTYLMWETS-----GVEIPPGIVNKELIIFGNQOEYEFHNPIFLKGL	1300	
Qy	455	TDY-----CF-----DY--NPDPENQIVGHQVILYL-----CHGMGQNOF	487	
Db	1301	EKYEQLPEDVGHCFTWADKFWQVYTYCKNKPDSQTILILEHAGSYFDEIQQRHGLA-NSI	1359	
Qy	488	FETYSQKEIRYNTHQ-----PEGCIAVEAGMDTLI-----MHLC-----BE	523	
Db	1360	SSYLIRPVQIRITKYQLLKELLTCCBEGKEIKDGLVNLSPKPRANDAMHLSMLEGFDE	1419	
Qy	524	TAPENOKFIIQEDGSLFHEQSKKCVQAARK-----ESSDSFVPLLRDCTNSDHOKWFFK	577	
Db	1420	NIESQELILQESFQVW--DPKTLIRKGRERHLFIPEMSLVFSKEVKD--SSGRSKILYK	1475	
Qy	578	ERM 580		
Db	1476	SKL 1478		
RESULT 13				
GGPP_NEUCR				
ID	GGPP_NEUCR	STANDARD;	PRT;	428 AA.
AC	P24322;			
DT	01-MAR-1992 (Rel. 21, Created)			
DT	01-MAR-1992 (Rel. 21, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Geranylgeranyl pyrophosphate synthetase (GGPP synthetase) (GGPPSASE)			
DE	(Geranylgeranyl diphosphate synthase) (Albino-3 protein) [includes:			
DE	Dimethylallyltransferase (EC 2.5.1.1); Geranyltransferase			
DE	(EC 2.5.1.10); Farnesyltransferase (EC 2.5.1.29)].			
GN	ALU-3 OR B8PB.010.			
OS	Neurospora crassa.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.			
OX	NCBI_TaxID=5141;			
XP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=74-OR23-1A / FGSC 987;			
RC	MEDLINE=91170267; PubMed=1826006;			


```
DR PRINTS: PR00134; GLHYDRASE10.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00591; GLYCOSYL HYDROL F10; 1.
DR PROSITE; PS0231; RICIN B LECTIN; 1.
KW xylan degradation; Hydrolase; Glycosidase; Signal; Lectin;
KW 3D-structure.
FT SIGNAL 1 41
FT CHAIN 42 477
FT DOMAIN 361 477
FT ACT_SITE 169 169
FT ACT_SITE 277 277
FT ACT_SITE 277 277
SQ SEQUENCE 477 AA; 51162 MW; E14A7PE37BDC68CC CRC64;

Query Match 3.5%; Score 110; DB 1; Length 477;
Best Local Similarity 19.8%; Pred. No. 0.68;
Matches 80; Conservative 60; Mismatches 165; Indels 100; Gaps 18;

Qy 220 RGVLTFLDCHCEHGLELQRIHEESAVVCPVID-----VIDNTFEY 267
Dy 120 RGHFLAW---HSQ-QPGMQL-----SGSALRQAMIDHNGVMAHYKGIQVQMDVYNE 169
Qy 268 L---GNSG-----EPOIGGPDW-----RLVFTHTVPERERIRMQSPVDV 304
Dy 170 APAGSSGARDSNLQSGNDWIEVAFRTARAADPSAKLCYNDYNVENWTWAKTQAMYNM 229
Qy 305 IRS-----PTMAGGLFAVSKKYFEYLGSDYTGMEVWNGENLEFSPRIWQCGVLETHPC 358
Dy 230 VRDFKQKQVPIDCVGFQSHFNSGSPYNSNFRITLQNFPAALGVDAITELDIQGA---PA 285
Qy 359 SHVGHVFPKQAPYSRNKALANS-----VRAAEVWMDFEKELYHNRPRALPEFGD 409
Dy 286 STYANV-----TNDCLAVSRCLGITVWGVDRSDSWSRSTQTLLEN-NDGSKAAAYTA 336
Qy 410 VTERKQLRDLQCKDFKWFLETVVPYVPELVHVPEDRPGFGLQNGKLTGYCFDYNPPDENQI 469
Dy 337 VL-----DALNGD-----SSEPADGQIQVGVG-SRGCLDV--PDASTS 373
Qy 470 VGHQVILYCHGMQGNQFPEYTSQKEIRYNTHQEGCIAVEAGMDTLIMHLCETAPENQ 529
Dy 374 DGTQLQLWDCHS-GTNQQAATDAGELRV---YGDKCLDAAGTSGKSVQIYSCWGGDNQ 429
Qy 530 KFILOEGSLFHEQSKCVQARKESSDSFVPLLRDCTNSDHQKW 574
Dy 430 KWLRLNSDGVVGVQSGCLDVAVGNGTANGTILQIYTCNSGNSQNRW 474

RESULT 15
RFBC_MYXA
ID RFBC_MYXA STANDARD; PRT; 1275 AA.
AC Q50864;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE O-antigen biosynthesis protein rfbC.
GN RFBC.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cyetobacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DK65640;
RA MEDLINE=96198166; PubMed=8626291;
RA Guo D., Bowden M.G., Pereshad R., Kaplan H.B.;
RT "The Myxococcus xanthus rfbABC operon encodes an ATP-binding cassette
RT transporter homolog required for O-antigen biosynthesis and
RL multicellular development."
RL J. Bacteriol. 178:1631-1639(1996).
CC -!- FUNCTION: INVOLVED IN O-ANTIGEN BIOSYNTHESIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.eib-sib.ch/announce/
CC or send an email to license@eib-sib.ch).
CC -----
DR EMBL; U36795; AAB05019.1; -.
DR PIR; T18556; T18556.
DR InterPro; IPR001296; Glyco_trans 1.
DR InterPro; IPR001173; Glyco_trans 2.
DR Pfam; PF00534; Glycos transf 1; 1.
DR Pfam; PF00535; Glycos transf 2; 2.
KW Lipopolysaccharide biosynthesis.
SQ SEQUENCE 1275 AA; 139596 MW; 3AF9662A10A140F1 CRC64;

Query Match 3.5%; Score 109.5; DB 1; Length 1275;
Best Local Similarity 21.9%; Pred. No. 2.9;
Matches 123; Conservative 62; Mismatches 194; Indels 183; Gaps 32;

Qy 43 AGAGAAEPGPPTPR-----PGRREPMPRPVPANALGARGEARVL--QLQGEELRLQE 95
Dy 214 AAAGAAQAGEAARRHVADVEALREPLAGR-TLPG-----ALRVTPQLWGEVLRQS 264
Qy 96 ----ESVRLHQINILSDRISLHRLPERWNPCKEKKYD----- 132
Dy 265 RFNAESV-----LALANLLGTRTAPPQPPSLADYPAMCAARESPARVTAQAEV 312
Qy 133 ---DNLPRTSVII-----AFYNEAWSTLRTVYSVLETSPIILLEEVILVDDYSREH 182
Dy 313 ARLSRRLISLVTPVHDASEAFIRECLASVSSQVYADWEW-----LLVDDASTAPH 363
Qy 183 LKERLANELSGLPKVLIRANKREGLVARLLGASAAARGDVLTPLDCHCEHGWLEP-- 240
Dy 364 LARILREAAERESRIRVLTASSEGDTARATNEGFAACRGDFVGFGLA---EDTLSPHA 418
Qy 241 ----LQRIHEESAVVCPVIDVIDWNTFYLGNSEPOIGGP---DWRLVFTWHTVPER 293
Dy 419 LAEVALAFLAQPELALLYTDDEGLDAQ-----GHRSAF-----FFKPDW----- 457
Qy 294 ERIRMQSPVDVIRSPMTMAGGLFAVSKKYFEYLGSDYTGMEVWNGENLEFSPRIWQCGV 353
Dy 458 -----SP-DLLRSVDYVRHFLVVRRETQAQVGLREGPD--GAQGHDLMLRL----- 501
Qy 354 ETHPCSHVGHVFP-----KQAPYSRNKALANSVRA-----AEVWMD- 389
Dy 502 -SEATSSIGHITPELVHAREGSAASASRGAGLDTATKAGVRLSEHLARQGESAEVTS 560
Qy 390 --EFKELYVHR-NPRAR-LEPPGDVTE-RKQLRDKLOCKDFKWFLETVVPYVPELVH- 439
Dy 561 PIQVRYVRYVRYGTPKVSIIIVPFKDRPDLLRLTLDL-----LAQTRYPHFVLLVSNN 613
Qy 440 EDRPGFGLQNG-----KGLTDYCFDYNPPDEN-----QIVGHQVILYCHGMQGNQ 488
Dy 614 STPETPFDLDEQWVDPRLVKTWHDHPNY-PAINNMAKQASG-ELLFLFLNDM---EVV 668
Qy 489 EYTSQKEIRYNTHQPE----GC 506
Dy 669 DPSWLDLVSAQORPEVGAAGC 690

Search completed: December 4, 2003, 19:42:50
Job time : 19 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 19:40:03 ; Search time 39 Seconds
(without alignments)
3844.322 Million cell updates/sec

Title: US-10-074-527-2

Perfect score: 3124

Sequence: 1 MWGRTARRRCPRELRRGREA.....LLRDCTNSDQKWFKKRML 581

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_xvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3124	100.0	581	4 Q8IXK2	Q8ixk2 homo sapien
2	3117	99.8	581	4 Q8NG54	Q8ng54 homo sapien
3	2708.5	86.7	576	11 Q8BGT9	Q8bgt9 mus musculus
4	1675	52.8	578	11 Q8H832	Q8h832 mus musculus
5	1651	52.8	578	4 Q8N4A0	Q8n4a0 homo sapien
6	1638	52.4	578	4 Q00208	Q00208 homo sapien
7	1562	50.0	284	4 Q9H771	Q9h771 homo sapien
8	1502	48.1	272	4 Q96CT9	Q96ct9 homo sapien
9	1198.5	38.4	623	5 Q95ZJ2	Q95zj2 caenorhabdi
10	1194.5	38.2	623	5 Q61391	Q61391 caenorhabdi
11	1190.5	38.1	622	11 Q8C7U7	Q8c7u7 mus musculus
12	1187.5	38.0	608	4 Q8IYH4	Q8iyh4 homo sapien
13	1185.5	37.9	622	4 Q8NC14	Q8nc14 homo sapien
14	1182.5	37.9	622	4 Q9UIV5	Q9uiv5 homo sapien
15	1180	37.8	626	5 Q9U2J8	Q9u2j8 caenorhabdi
16	1177	37.7	624	5 Q95ZJ1	Q95zj1 caenorhabdi

17	1176.5	37.7	622	11 Q8CED2	Q8ced2 mus musculus
18	1176	37.6	626	5 Q61392	Q61392 caenorhabdi
19	1173	37.5	624	5 Q61393	Q61393 caenorhabdi
20	1156.5	37.0	633	11 P70419	P70419 mus musculus
21	1142.5	36.6	633	4 Q14435	Q14435 homo sapien
22	1126	36.0	650	5 Q8MRC9	Q8mrc9 drosophila
23	1126	36.0	650	5 Q9V7T0	Q9v7t0 drosophila
24	1123.5	36.0	556	11 Q8BLE4	Q8ble4 mus musculus
25	1122.5	35.9	516	4 Q96PX0	Q96px0 homo sapien
26	1122.5	35.9	556	4 Q8IUC8	Q8iuc8 homo sapien
27	1122.5	35.9	556	11 Q8CF93	Q8cf93 mus musculus
28	1102	35.3	623	11 Q9QZ16	Q9qz16 mus musculus
29	1094.5	35.0	552	11 Q8BRT3	Q8bvt3 mus musculus
30	1089.5	34.9	536	5 Q95T43	Q95t43 drosophila
31	1089.5	34.9	630	5 Q9VMU3	Q9vmu3 drosophila
32	1023	32.7	639	4 Q96C46	Q96c46 homo sapien
33	1023	32.7	639	4 Q8N3T1	Q8n3t1 homo sapien
34	1023	32.7	639	4 Q96DJ5	Q96dj5 homo sapien
35	1013.5	32.4	571	4 Q10471	Q10471 homo sapien
36	1001	32.0	570	11 Q99ME1	Q99me1 mus musculus
37	999	32.0	589	5 Q61390	Q61390 caenorhabdi
38	997	31.9	443	5 Q9VQ02	Q9vq02 caenorhabdi
39	997	31.9	443	5 Q9VQ02	Q9vq02 drosophila
40	992	31.8	603	11 Q925R7	Q925r7 rattus norv
41	991.5	31.7	603	4 Q8IXJ2	Q8ixj2 homo sapien
42	991	31.7	570	11 Q8BL27	Q8bl27 mus musculus
43	990	31.7	526	11 Q922K5	Q922k5 mus musculus
44	989.5	31.7	579	5 Q61398	Q61398 caenorhabdi
45	988.5	31.6	579	5 Q9U2C4	Q9u2c4 caenorhabdi

ALIGNMENTS

RESULT 1

Q8IXK2 PRELIMINARY; PRT; 581 AA.

AC Q8IXK2; 01-MAR-2003 (Tremblrel. 23, Created)

DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE UDP-GalNAC-transferase 12.

GN GALNT12.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Bennett E.P.;

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ132365; CAC80100.2; -

KW Transferase.

SQ SEQUENCE 581 AA; 66938 MW; 8C001D58E103A523 CRC64;

Query Match	100.0%;	Score	3124;	DB	4;	Length	581;
Best Local Similarity	100.0%;	Pred. No.	3.6e-246;				
Matches	581;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MWGRTARRRCPRELRRGREALLVLLALLAGLGSVLRAQRGAGAGAAEPGPPRTPRGR	60				
Db	1	MWGRTARRRCPRELRRGREALLVLLALLAGLGSVLRAQRGAGAGAAEPGPPRTPRGR	60				
Qy	61	REVPMPRPVPNALGARGAENVLQGEELRQEEVRLHQNIVLSDRISLHRLRPER	120				
Db	61	REVPMPRPVPNALGARGAENVLQGEELRQEEVRLHQNIVLSDRISLHRLRPER	120				
Qy	121	WNFLCKEKKYDYNLPRTSVIIAFYNEAWSTLLRTVYVLETSPTDILLEEVILVDDYSR	180				
Db	121	WNFLCKEKKYDYNLPRTSVIIAFYNEAWSTLLRTVYVLETSPTDILLEEVILVDDYSR	180				
Qy	181	EHLKERLANELSGLPKVRIFRANKREGLVRALLGSAARGDVLTFDCHCEHGLEP	240				
Db	181	EHLKERLANELSGLPKVRIFRANKREGLVRALLGSAARGDVLTFDCHCEHGLEP	240				

QY 181 EHLKERLANELSGLPKVLRLTRANKREGVLRARLLGASAGRDVLTFLDCHCECHGWLEP 240
DB 176 EHLKERLANELSGLPKVLRLTRANKREGVLRARLLGASAGRDVLTFLDCHCECHGWLEP 235
QY 241 LLQRIHEEESAVCPVLDVNDWNTFEYLGNSGEPOIGGFDWRLVFTWHTVPERIRMQS 300
DB 236 LLQRIHEEESAVCPVLDVNDWNTFEYLGNSGEPOIGGFDWRLVFTWHTVPERIRMQS 295
QY 301 PVDVIRPTWAGGLFAVSKYFEYLGSDYDTGMEVWGGENLEFSPRIWCCGVLETHPCSH 360
DB 296 PIDVIRPTWAGGLFAVSKYFEYLGSDYDTGMEVWGGENLEFSPRIWCCGVLETHPCSH 355
QY 361 VGHVFPKQAPYSRNKALANSVRAAEVWDEPKELYIYRNPRARLEPFGDVTERRKQLRDLK 420
DB 356 VGHVFPKQAPYSRNKALANSVRAAEVWDEPKELYIYRNPRARLEPFGDVTERRKQLRDLK 415
QY 421 QCKDFKWFLETVYVPELVHVPEDRPGFGMLQNGKLTDCYFNDPNDENQIVGHQVILVLYCH 480
DB 416 QCKDFKWFLETVYVPELVHVPEDRPGFGMLQNGKLTDCYFNDPNDENQIVGHQVILVLYCH 475
QY 481 GMGNQFPEYTSQKEIRYNTHQPEGCIAVEAGMDTLIMHLCEETAPENQKFIQEDGSLF 540
DB 476 GMGNQFPEYTSQKEIRYNTHQPEGCIAVEAGMDTLIMHLCEETAPENQKFIQEDGSLF 535
QY 541 HGQSKCVQAKRESDSPVPLLRDCTNSDHQKFFKERM 580
DB 536 HGQSKCVQAKRESDSPVPLLRDCTNSDHQKFFKERM 575

RESULT 4

O08832 ID O08832 PRELIMINARY; PRT; 578 AA.
AC O08832;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) (Protein-
DE UDP acetylglucosaminyltransferase)
DE acetylglucosaminyltransferase)
DE acetylglucosaminyltransferase)
GN GALNT4 OR PGCANFASB-T4.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RX MEDLINE=97298094; PubMed=9153242;
RA Hagen F.K., Ten Hagen K.G., Beres T.M., Balys M.M.,
RA VanWuyckhuysen B.C., Tabak L.A.;
RT "CDNA cloning and expression of a novel UDP-N-acetyl-D-
RT galactosamine:polypeptide N-acetylglucosaminyltransferase.";
RL J. Biol. Chem. 272:13843-13848(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
CC -1- FUNCTION: THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED
CC OLIGOSACCHARIDE BIOSYNTHESIS. THE TRANSFER OF AN N-ACETYL-D-
CC GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE
CC PROTEIN RECEPTOR. AGT AND PAG4 HAVE DIFFERENT SUBSTRATES
CC SPECIFICITY.
CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + POLYPEPTIDE =
CC UDP + N-ACETYL-D-GALACTOSAMINYL-POLYPEPTIDE.
CC -1- PATHWAY: GLYCOSYLATION.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC -1- TISSUE SPECIFICITY: HIGHER EXPRESSION IN SUBLINGUAL GLAND,
CC STOMACH, SMALL INTESTINE, CERVIX AND COLON AND TO A LESSER EXTENT

CC IN KIDNEY, OVARY, LUNG AND UTERUS. LOW LEVELS IN SPLEEN AND TRACE
CC LEVELS IN LIVER, HEART AND BRAIN. NO EXPRESSION IN SUBMANDIBULAR
CC AND PAROTID GLANDS, SKELETAL MUSCLE AND TESTIS.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
DR EMBL; U73819; AAB58301.1; -;
DR EMBL; AK033494; BAC28317.1; -;
DR MGI; 894692; Galnt4.
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00535; Glycos_transf_2; 1.
DR Pfam; PF00652; Ricin_B_lectin; 2.
DR SMART; SM00458; RICIN_1.
DR PROSITE; PS00231; RICIN_B_LECTIN; 1.
KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW Golgi stack; Glycoprotein.
FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 13 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 32 578 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 578 AA; 66554 MW; 10ADC0DB8B30835 CRC64;
Query Match 53.6%; Score 1675; DB 11; Length 578;
Best Local Similarity 57.9%; Pred. No. 5.7e-128;
Matches 330; Conservative 78; Mismatches 144; Indels 18; Gaps 8;
QY 23 VLIALALAGL-----GSVLRAQAGAGAAEPGPRTPRPGRRE-----PVMPRPVPA 72
DB 13 LLLALLTLAVILVEFVSSTLYASPGAG-GARELGPRRLPDLDTREEDLSQLYIKPPADS 71
QY 73 NALGARGEAVRLQOGEELRLOEESVRLHQINILYLSDRISLHRLPERWNPCLKEKYDY 132
DB 72 HALGEWGRASKLQNEGELKQKEELIERYAINILYLSDRISLHRIEDKRYEYCKAKFY 131
QY 133 DNLPRTSVIAFYNEAWSTLLRTVSVLETSPDILLEEVILVDVDSYDREHLKERLANELS 192
DB 132 RSLPTTSVIAFYNEAWSTLLRTVSVLETSPDILLEEVILVDVDSYDREHLKERLANELS 191
QY 193 GLPKVRLIRANKREGVLRARLLGASAGRDVLTFLDCHCECHGWLEPLQRIHEESAV 252
DB 192 NLERVLIRTNKREGVLRARLLGATFATGDLVTLFDCHECNCNTGWLEPLERISRDSTA 251
QY 253 VCPVIDVIDWNTFEYLGNSGEPOIGGFDWRLVFTWHTVPERIRMQSPVDVIRSPMAG 312
DB 252 VCPVIDVIDWNTFEYLGNSGEPOIGGFDWRLVFTWHTVPERIRMQSPVDVIRSPMAG 311
QY 313 GLFAVSKYFEYLGSDYDTGMEVWGGENLEFSPRIWCCGVLETHPCSHVGHVFPKQAPYS 372
DB 312 GLFAVSKYFEYLGSDYDTGMEVWGGENLEFSPRIWCCGVLETHPCSHVGHVFPKQAPYS 371
QY 373 RNKALANSVRAAEVWDEPKELYIYRNPRARLEPFGDVTERRKQLRDLKDFKWFLETV 432
DB 372 RPNFLQNTARAAEVWDEPKELYIYRNPRARLEPFGDVTERRKQLRDLKDFKWFLETV 431
QY 433 YPELVHVPEDRPGFGMLQNGKLTDCYFNDPNDENQIVGHQVILVLYCHGQVQFFEYTS 492
DB 432 FSNLHVPEDRPGFGMLQNGKLTDCYFNDPNDENQIVGHQVILVLYCHGQVQFFEYTS 490
QY 493 QKEIRYNTHQPEGCIAVEAGMDTLIMHLCEE---TAPENQKFIQEDGSLFHSQSKCVQ 549
DB 491 NKIRFNS-VTELCAEVPOOKDVTGQNCNCPKGLPVPVNIHFKEDGTIFHPHTRCLLS 549
QY 550 AARKESDSDVPVPLLRDCTNSD-HQKFFKFE 578
DB 550 AYRTAGRSPSVH-MKTCDALDKQLWRFK 578
RESULT 5
Q8N4A0 ID Q8N4A0 PRELIMINARY; PRT; 578 AA.
AC Q8N4A0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)


```
01-MAR-2001 (TrEMBLrel. 16, Created)
01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ21212.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Ohtani R., Ota T., Suzuki Y., Ohayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isegai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK024865; BAB15027.1; -.
DR InterPro; IPR000772; Ricin B lectin.
DR Pfam; PF00652; Ricin B lectin; 2.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
KW Hypothetical protein.
SQ SEQUENCE 284 AA; 33037 MW; 3DD5FD6F1154AB71 CRC64;

Query Match 50.0%; Score 1562; DB 4; Length 284;
Best Local Similarity 100.0%; Pred. No. 3.5e-119;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 MQSPVDVIRPTMAGGLPAVSKKYFEYLGSDYTGMEVWGGENLEFSFRIMQCGGVLETHP 357
DB 1 MQSPVDVIRPTMAGGLPAVSKKYFEYLGSDYTGMEVWGGENLEFSFRIMQCGGVLETHP 60

QY 358 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHNRNPRARLEPFGDVTERKQLR 417
DB 61 CSHVGHVFPKQAPYSRNKALANSVRAAEVWMDPEFKELYHNRNPRARLEPFGDVTERKQLR 120

QY 418 DKLOKDFKFLVTVYPPELHVPDRPGFPGMLQNGKLTDCYFDYNDPNDENQIVGHQVILY 477
DB 121 DKLOKDFKFLVTVYPPELHVPDRPGFPGMLQNGKLTDCYFDYNDPNDENQIVGHQVILY 180

QY 478 LCHGMGNQFPEYTSQKEIRYNTHQPEGCIAVEAGMDTLIMHLCETAPENQKFILOEDG 537
DB 181 LCHGMGNQFPEYTSQKEIRYNTHQPEGCIAVEAGMDTLIMHLCETAPENQKFILOEDG 240

QY 538 SLFHEQSKKCVQARKSSDSFVPLLRDCTNSDHQKWFKERM 581
DB 241 SLFHEQSKKCVQARKSSDSFVPLLRDCTNSDHQKWFKERM 284

RESULT 8
Q96CT9 ID Q96CT9 PRELIMINARY; PRT; 272 AA.
AC Q96CT9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Similar to hypothetical protein FLJ21212.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RA Straube B.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013945; AAH13945.1; -.
DR InterPro; IPR000772; Ricin B lectin.
DR Pfam; PF00652; Ricin B lectin; 2.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
KW Hypothetical protein.
SQ SEQUENCE 272 AA; 31726 MW; 6E78ADD3FC84390F CRC64;
```

```
Query Match 48.1%; Score 1502; DB 4; Length 272;
Best Local Similarity 100.0%; Pred. No. 2.6e-114;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 MAGGLFAVSKKYFEYLGSDYTGMEVWGGENLEFSFRIMQCGGVLETHPCSHVGHVFPKQA 369
DB 1 MAGGLFAVSKKYFEYLGSDYTGMEVWGGENLEFSFRIMQCGGVLETHPCSHVGHVFPKQA 60

QY 370 PYSRNKAANSVRAAEVWMDPEFKELYHNRNPRARLEPFGDVTERKQLRDKLOKDFKFL 429
DB 61 PYSRNKAANSVRAAEVWMDPEFKELYHNRNPRARLEPFGDVTERKQLRDKLOKDFKFL 120

QY 430 ETVPPELHVPDRPGFPGMLQNGKLTDCYFDYNDPNDENQIVGHQVILYLCGMGNQFPE 489
DB 121 ETVPPELHVPDRPGFPGMLQNGKLTDCYFDYNDPNDENQIVGHQVILYLCGMGNQFPE 180

QY 490 YTSQKEIRYNTHQPEGCIAVEAGMDTLIMHLCETAPENQKFILOEDGSLFHEQSKKCVQ 549
DB 181 YTSQKEIRYNTHQPEGCIAVEAGMDTLIMHLCETAPENQKFILOEDGSLFHEQSKKCVQ 240

QY 550 AARKSSDSFVPLLRDCTNSDHQKWFKERM 581
DB 241 AARKSSDSFVPLLRDCTNSDHQKWFKERM 272
```

```
RESULT 9
Q95ZJ2 ID Q95ZJ2 PRELIMINARY; PRT; 623 AA.
AC Q95ZJ2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Y39E4B.12b protein.
GN Y39E4B.12 OR Y39E4B.12b.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Barlow K.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL110487; CAC42368.1; -.
DR WormPep; Y39E4B.12b; CE28119.
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00535; Glycos transf_2; 1.
DR Pfam; PF00652; Ricin B lectin; 3.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS00630; IMP_2; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
SQ SEQUENCE 623 AA; 71013 MW; 722AC7B93EF5FB4D CRC64;
```

```
Query Match 38.4%; Score 1198.5; DB 5; Length 623;
Best Local Similarity 46.0%; Pred. No. 4.6e-89;
Matches 256; Conservative 74; Mismatches 178; Indels 51; Gaps 17;

QY 51 GPPRTPRP-----GRREP-----VMPRPV-----PANALGARGEAV---RLQ 85
DB 66 GPPIEPEPVVNNKVEEQPGGLAKPKFVDPNDPIYKKGDAQAQAGELGKAVVVDKTK 125

QY 86 LQGEELRQESVRLHQINILYSDRISLHRLRLPERWNPLCKEKKYDYNLPRTSVIAFY 145
DB 126 LSTEERAKYDKGMLNNAFNQYASDMISVHRTLPTNIDAECKTEKYN-ENLPRTSVIICFH 184
```

QY 146 NEAWSTLLRTVYSVLETSPTDILLLEEVILVDDYSDREHLKERLANELSGL-PKVLIRANK 204
 DB 185 NEAWSVLLRTVHSVLERTPDHLLLEEVILVDDYSDREHLKERLANELSGL-PKVLIRANK 244
 QY 205 REGIUVARLLGASAAAGDVLTFDCHCECHGWLPLELLORIHEESAVVCPVIDIDWNT 264
 DB 245 REGLIRARLGAATAVATGEVLTYLDSHCECHGWLPLELLORIHEESAVVCPVIDIDWNT 304
 QY 265 FEYLGNSGE-POIGGDFWRLVFTWHTVPERIRMQSPVDVIRSPMTAGGLFAVSKKYE 323
 DB 305 FEYHSHKAYFTSVGGDFWGLQFNWHSIPERDRKTRTPIDVRSPTWAGGLFSIDKKYE 364
 QY 324 YLGSYDTGMEVWGENLEFSFRIWQCGGVLETHPCSHVGHVFPKQAPY-SSRNKALAN 379
 DB 365 KLGTYPGDFDIWGENLELSFKIWMCGGTLEIVPCSHVGHVFPKQAPY-SSRNKALAN 424
 QY 380 SVRAAEVMDDEFKELYHNRNPRARLEPGDVTERRKQLDKLQCKDFKWFLETVPVPELHVP 439
 DB 425 SIRLAELWLDYKTYYYER-INNLQDGFDISRRKRLREDLGCKSFKWYLDNIYPELFPV 483
 QY 440 EDRPGFGLQNLKGLTDYCPDYNPPDENQIVGHQVILYLCHGMQONQFFETYSQKEIRYN 499
 DB 484 GESVAKGEMRNAGGNKQCIDYKPSG-----GKTVMGYQCHNOGNOYMWLSKDGEIR-- 536
 QY 500 THOPEGCIAVEAGMDTLIMHLCBETAPENOKFIQED-GSLFHEQSKKQVQAARKESSDS 558
 DB 537 --RDESCVDY-AGSDVMVFP-C-HGMKGNQEWRYNHTDTRGLQHAVSKCLGWTGDKAKLE 591
 QY 559 FVPLLRDCTNSD-HQKWFPE 578
 DB 592 MVA-----CQYDDPYQHWFPE 608

RESULT 10

Q61391 ID O61391 PRELIMINARY; PRT; 623 AA.
 AC O61391;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE GLY5a (EC 2.4.1.41).
 GN GLY-5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
 OC Rhabditiida; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=N2;
 RX MEDLINE=98192620; PubMed=9525933;
 RA Hagen F.K., Nehrke K.;
 RT "cDNA cloning and expression of a family of UDP-N-acetyl-
 RT Dgalactosamine:Polypeptide N-acetylglucosaminyltransferase sequence
 RT homologs from Caenorhabditis elegans";
 RL J. Biol. Chem. 273:8268-8277 (1998).
 DR EMBL; AF031835; AAC33671.1; --
 DR InterPro; IPR001173; Glyco.trans.2.
 DR InterPro; IPR000760; Inositol P.
 DR InterPro; IPR000772; Ricin B lectin.
 DR Pfam; PF00535; Glycos.transf.2; 1.
 DR Pfam; PF00652; Ricin B lectin; 3.
 DR SMART; SM00458; RICIN; 1.
 DR PROSITE; PS00630; IMP2; 1.
 DR PROSITE; PS0231; RICIN B LECTIN; 1.
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 623 AA; 71014 MW; 9C2AC05F9EF5FC6D CRC64;

Query Match 38.2%; Score 1194.5; DB 5; Length 623;
 Best Local Similarity 45.8%; Pred. No. 9.8e-89;
 Matches 257; Conservative 75; Mismatches 178; Indels 51; Gaps 17;

QY 51 GPPRTPR-----GRREP-----VMPPPPV-----PANALGARGEAV---RLQ 85

DB 66 GPPIEPEPVVNNKVBEEEPQCGNLAKPKFMDPNDPIYKKGDAQAAGELGKAVVVDKTK 125
 QY 86 LQGEELRLQEEVRLHQINLYSLRHLRLPERWNLCKEKKYDNDLPRTSVITAFY 145
 DB 126 LSTEEKAKYDGMNNAFOYASMTLPTNIDAECKTEKYN-ENLPRTSVITAFY 184
 QY 146 NEAWSTLLRTVYSVLETSPTDILLLEEVILVDDYSDREHLKERLANELSGL-PKVLIRANK 204
 DB 185 NEAWSVLLRTVHSVLERTPDHLLLEEVILVDDYSDREHLKERLANELSGL-PKVLIRANK 244
 QY 205 REGIUVARLLGASAAAGDVLTFDCHCECHGWLPLELLORIHEESAVVCPVIDIDWNT 264
 DB 245 REGLIRARLGAATAVATGEVLTYLDSHCECHGWLPLELLORIHEESAVVCPVIDIDWNT 304
 QY 265 FEYLGNSGE-POIGGDFWRLVFTWHTVPERIRMQSPVDVIRSPMTAGGLFAVSKKYE 323
 DB 305 FEYHSHKAYFTSVGGDFWGLQFNWHSIPERDRKTRTPIDVRSPTWAGGLFSIDKKYE 364
 QY 324 YLGSYDTGMEVWGENLEFSFRIWQCGGVLETHPCSHVGHVFPKQAPY-SSRNKALAN 379
 DB 365 KLGTYPGDFDIWGENLELSFKIWMCGGTLEIVPCSHVGHVFPKQAPY-SSRNKALAN 424
 QY 380 SVRAAEVMDDEFKELYHNRNPRARLEPGDVTERRKQLDKLQCKDFKWFLETVPVPELHVP 439
 DB 425 SIRLAELWLDYKTYYYER-INNLQDGFDISRRKRLREDLGCKSFKWYLDNIYPELFPV 483
 QY 440 EDRPGFGLQNLKGLTDYCPDYNPPDENQIVGHQVILYLCHGMQONQFFETYSQKEIRYN 499
 DB 484 GESVAKGEMRNAGGNKQCIDYKPSG-----GKTVMGYQCHNOGNOYMWLSKDGEIR-- 536
 QY 500 THOPEGCIAVEAGMDTLIMHLCBETAPENOKFIQED-GSLFHEQSKKQVQAARKESSDS 558
 DB 537 --RDESCVDY-AGSDVMVFP-C-HGMKGNQEWRYNHTDTRGLQHAVSKCLGWTGDKAKLE 591
 QY 559 FVPLLRDCTNSD-HQKWFPE 578
 DB 592 MVA-----CQYDDPYQHWFPE 608

RESULT 11

Q8C7U7 ID Q8C7U7 PRELIMINARY; PRT; 622 AA.
 AC Q8C7U7;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE UDP-N-acetyl-alpha-D-galactosamine:polypep tide
 DE N-acetylglucosaminyltransferase 6.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 DR EMBL; AK049222; BAC33618.1; --
 SQ SEQUENCE 622 AA; 71537 MW; 6DBC3A3A2CCPDBDAA CRC64;

Query Match 38.1%; Score 1190.5; DB 11; Length 622;
 Best Local Similarity 44.5%; Pred. No. 2.1e-88;
 Matches 249; Conservative 96; Mismatches 169; Indels 45; Gaps 13;

QY 50 GPPRT-----PRGRREPVPMPPPANALGARGEAV-RLQGEELRLQEE 96
 DB 76 PEPPTLVSTVHNSCLPGFYTPAELKPFWDPPDPNSPGADGKAFQKKEWTNLTKEKE 135

```

QY 97 SVRLHQNIVLSRISLHRL-PERWNPLCCKEKKY-DYDNLPRTSVLIIFNEAWSTLLR 154
Db 136 GYKXKHCENAFASORISLQSLGDPDTRPEPCDFRCPLPTTSVLIIFHNEAWSTLLR 195
QY 155 TVSVLETSPOILLEEVILVDDSDRHLKERLANELSGLPKVRILIRANKREGILVARLL 214
Db 196 TVSVLHTSPAILLKEILLVDDASTDEHLKERLEQYVQQLQIVRVQRERKGLITARLL 255
QY 215 GASARGDVLTFLDCHCEHEGMLPLQLRIHEESAVVCPVIDVDMNTFEY---LGNS 271
Db 256 GASVAQAEVLTFLDAHCECFHGLPLLARIAEDKTAVVSPDIVTIDLNTFQSRPVRG 315
QY 272 GEPOIGGFDRVLTWHVTPERIRMQSPVDVIRSPMTAGGLPAVSKKYFEYLGSDVTG 331
Db 316 KAHSRGNFWSLFGWEMLPEHEKQRKDETYPIKSPTFAGGLFSISKAYFEHIGTYDNQ 375
QY 332 MEVWGENLEFSFRIWOCGVLETHPCSHGVHPFKOAPYSRNKALA---NSVRAAEVW 387
Db 376 MEIWGENVEMSRVWQCGQLIIFCSVGVHVFRTKSPHTFPKGTSVIARNQVRLAEVW 435
QY 388 MDEFKELYHRNPRA---RLEPFGDVTERKQLDKLQCKDFKWFLETVYPPELHVPEDRP 443
Db 436 MDDYKIFFYRNLOQAAMQWQENFGDISERLRLREQLRCHNFSWYLNHNVPFVPELNP 495
QY 444 GFTGMLQNKGLTDYCDYNPPDENQIVGHQVILYLCHGMQONOFFEYTSOKERYNTHQ 503
Db 496 TPFYGAIKNLG-TNOCILDV---ENNRGGKPLIMVCHNLGNQYFEYTSQDLRHNIQ 551
QY 504 EGCIAVEAGMTLIMLCE---ETAPENOKFILOEDGSLFHEQSKKVOAARKSSDS 558
Db 552 ---LCHASGSLGLASCQFVGNRSVPKDEEWELTQDLIRNSGGTCLTSQDKK---- 604
QY 559 FVPLLRDCTNSD-HQKWF 576
Db 605 --PAMAPCNPRDPQLWLF 621

RESULT 12
Q81YH4
ID Q81YH4 PRELIMINARY; PRT; 608 AA.
AC Q81YH4;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DE UDP-N-acetyl-alpha-D-galactosamine:polypeptide
DE N-acetylglucosaminyltransferase 6 (GalNAc-T6).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Isogai T., Ota T., Nishikawa T., Hayaishi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074658; BAC11118.1; --
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR000772; Ricin B lectin.
DR Pfam; PF00535; Glycos transf 2; 1.
DR SMART; SM00652; Ricin_B lectin; 1.
DR SMART; SM00458; RICIN_1.
DR PROSITE; PS50231; RICIN_B_LLECTIN; 1.
KW Hypothetical protein.
SQ SEQUENCE 608 AA; 69357 MW; A922107DA2E243F CRC64;

Query Match 38.0%; Score 1187.5; DB 4; Length 608;
Best Local Similarity 45.9%; Pred. No. 3.5e-88;
Matches 249; Conservative 91; Mismatches 163; Indels 39; Gaps 14;

QY 58 PGRREVPMPRPVPANALGARGEAVR----LQIQGEELRLQESVRLHQINIVLSRISL 113
Db 82 PAELKPFWRPPQDPNAPGADGKAFQSKWTPLEQ---KEGYKXHCENAFASRISL 138
QY 114 HRL-PERWNPLCCKEKKY-DYDNLPRTSVLIIFNEAWSTLLRTVSVLETSPOILLEEV 171
Db 139 QRSGLGDPTRPPECVDQKFRCPPLATTSVLIIFHNEAWSTLLRTVSVLHTTFAILLKEI 198
QY 172 ILVDDYSDRHLKERLANELSGLPKVRILIRANKREGILVARLLGASARGDVLTFLDCHC 231
Db 199 ILVDDASTEEHLKERLEQYVQQLQIVRVVRQERKGLITARLLGASVAQAEVLTFLDAH 258

```

```

QY 232 ECEHGLEPLQLRIHEESAVVCPVIDVDMNTFEYLGNSGEPI---GGFDWRLVFTWH 288
Db 259 ECFHGLEPLLARIAEDKTAVVSPDIVTIDLNTFEFAKPVQGRVHSRGNFDSLTFGE 318
QY 289 TVPERIRMQSPVDVIRSPMTAGGLPAVSKKYFEYLGSDVTGMEVWGENLEFSFRIW 348
Db 319 TLPHEKQRKDETYPIKSPTFAGGLFSISKAYFEHIGTYDNQMEIHWGENVEMSRVW 378
QY 349 CGVLETHPCSHGVHPFKOAPYSRNKALA---NSVRAAEVWMDPEKELYHRN----P 400
Db 379 CGQLIIFCSVGVHVFRTKSPHTFPKGTSVIARNQVRLAEVWMSYKLI FYRNLOQA 438
QY 401 RARLEPFGDVTERKQLDKLQCKDFKWFLETVYPPELHVPEDRPFGFQMLQNKGLTDCFD 460
Db 439 MAQESFGDISERLRLREQLRCHNFSWYLNHNVPFVPELDTFTFYGAIKNLG-TNOCILD 497
QY 461 YNPDPENQIVGHQVILYLCHGMQONOFFEYTSOKERYNTHQEGCIAVEAGMTLIMHL 520
Db 498 VG---ENNRGGKPLIMYSGHLGNQYFEYTTQDLRHNIK-QLCLHVSKG--ALGLS 551
QY 521 CEET-----APENOKFILOEDGSLFHEQSKKVOAARKSSDSFVPLLRDCTNSD-HQKW 574
Db 552 CHTGKNSQVPKDEEWELAQDLIRNSGGTCLTSQDKK-----PAMAPCNPSDPHQLW 605
QY 575 FF 576
Db 606 LF 607

RESULT 13
Q8NCL4
ID Q8NCL4 PRELIMINARY; PRT; 622 AA.
AC Q8NCL4;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ90177.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Isogai T., Ota T., Nishikawa T., Hayaishi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074658; BAC11118.1; --
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR000772; Ricin B lectin.
DR Pfam; PF00535; Glycos transf 2; 1.
DR SMART; SM00652; Ricin_B lectin; 1.
DR SMART; SM00458; RICIN_1.
DR PROSITE; PS50231; RICIN_B_LLECTIN; 1.
KW Hypothetical protein.
SQ SEQUENCE 622 AA; 71216 MW; 9F839D35D0A597ED CRC64;

Query Match 37.9%; Score 1185.5; DB 4; Length 622;
Best Local Similarity 45.9%; Pred. No. 5.3e-88;
Matches 249; Conservative 91; Mismatches 163; Indels 39; Gaps 14;

QY 58 PGRREVPMPRPVPANALGARGEAVR----LQIQGEELRLQESVRLHQINIVLSRISL 113
Db 96 PAELKPFWRPPQDPNAPGADGKAFQSKWTPLEQ---KEGYKXHCENAFASRISL 152
QY 114 HRL-PERWNPLCCKEKKY-DYDNLPRTSVLIIFNEAWSTLLRTVSVLETSPOILLEEV 171
Db 153 QRSGLGDPTRPPECVDQKFRCPPLATTSVLIIFHNEAWSTLLRTVSVLHTTFAILLKEI 212

```

```

QY 172 ILVDDSDREHLKERLANELSGLPKVLRLRANKREGVLRLGASARGDVLTFDCHC 231
Db 213 ILVDDASTEHLKEKLEQYVYKQVVRVVRQBERKGLITARLLGASVAQAEVLTFDACH 272
QY 232 ECEHGLEPPLLORLHIEESAVVCPVIDVIDWNTFEYLGNSGEPIQI---GGFDWRLVFTWH 288
Db 273 ECFHGLEPPLLARIAEDKTIVVSPDIVTDLNTEFEPAKPVQGRVHSGRGNFDSLTFGWE 332
QY 289 TVPERERIMQSPVDVIRSPMTAGGLFAVSKYFEYLGSYDTGMEVWGGNLEFSFRIMQ 348
Db 333 TLPPEKQRKQKDETYPIKSPTEAGGLFSPKSYFEHIGTYDQMEIWGGENVEMSPRVQ 392
QY 349 CGGVLETHPCSHGVHPKQAPYSRNKALA-----NSVRAAEVWMDSEKELYHNRNPA-- 402
Db 393 CGGLEIIPCSVVGHVFTKSPHTFPKGTSVIARNQVRLAEVWMDSYKKIFYRNLQAOK 452
QY 403 --RLEPFGDVTERRKQDLKQDFKWFLETVVPELHVPEDRPGFFGMLQNGLTDCYCFD 460
Db 453 MTQEKSGDISERQLREQHLCHNFSWYLNHVPPEMFPVPLTFTFYGAIKNLG-TNQCLD 511
QY 461 YNPDPENQIVGHQVILYLCHGMGNQFFEYTSQKEIRYNTHQPEGCIABAGMDTLIMHL 520
Db 512 VG---ENNRGGKPLIMYSCHGLGNGQYFEYTTQDLRHNIAT-RLCLHVSKG--ALGLGS 565
QY 521 CEET-----APENQKFILOEDGSLFHEQSKKCVQAARKESSDFVPLLRDCTNSD-HOKW 574
Db 566 CFTGKNSQVPKDEWEALAQDLIRNSGSGTCLTSQDKK-----PAMAPCNFSDPHQLW 619
QY 575 FF 576
Db 620 LF 621

RESULT 14
QYUIV5 PRELIMINARY; PRT; 622 AA.
AC QYUIV5 (PRELIMINARY; PRT; 622 AA.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE UDP-GalNac:polypeptide N-acetylglucosaminyltransferase.
GN GALNAC-T6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gastric;
RX MEDLINE=9395099; PubMed=10464263;
RA Bennett E.P., Haas H., Mandel U., Hollingsworth M.A., Akisawa N.,
RA Ikematsu Y., Merckx G., Geurts van Kessel A., Olofsson S., Clausen H.,
RT "Cloning and Characterization of a Close Homologue of Human UDP-N-
RT acetyl--D-galactosamine:Polypeptide N-Acetylglucosaminyltransferase-
RT T3, Designated GalNac-T6. Evidence for genetic but not functional
RT redundancy."
RL J. Biol. Chem. 274:25362-25370(1999).
DR EMBL; Y08565; CAA69876.1; --
DR Genbank; HGNC:4128; GALN6.
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00535; Glycos_transf_1.
DR Pfam; PF00652; Ricin_B_lectin; 2.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS0231; RICIN_B_LECTIN; 1.
KW Transferase.
SQ SEQUENCE 622 AA; 71168 MW; 3944AF2AF0916EF CRC64;

Query Match 37.9%; Score 1182.5; DB 4; Length 622;
Best Local Similarity 45.8%; Pred. No. 9.3e-88;
Matches 248; Conservative 91; Mismatches 164; Indels 39; Gaps 14;

```

RESULT 15

```

QYUIV5 PRELIMINARY; PRT; 622 AA.
AC QYUIV5 (PRELIMINARY; PRT; 622 AA.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Y39E4B.12a protein.
GN Y39E4B.12 OR Y39E4B.12A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Barlow K.;
RA Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
DR none;
DR "Genome sequence of the nematode C.elegans: A platform for
DR investigating biology."
DR Science 282:2012-2018(1998).
DR EMBL; AL110487; CAB54435.1; --
DR WormPep; Y39E4B.12a; CB24240.
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR000760; Inositol_P_.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00535; Glycos_transf_2; 1.
DR Pfam; PF00652; Ricin_B_lectin; 3.
DR SMART; SM00458; RICIN; 1.

```

Search completed: December 4, 2003, 19:43:42
Job time : 41 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 19:38:37 ; Search time 47 Seconds
(without alignments)
1962.131 Million cell updates/sec

Title: US-10-074-527-2

Perfect score: 3124

Sequence: 1 MWGTTARRCPRELRRGREA.....LLRDCTNSDQKWFKKERML 581

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.+

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3124	100.0	581	23	ABG32510 Human glycosyltran
2	3124	100.0	581	23	AAE25019 Human drug metabol
3	1788	57.2	355	22	ABG04767 Novel human diagn
4	1638	52.4	578	24	ABU07527 Human N-acetylgl
5	1369	43.8	256	21	AAE53452 Human colon cancer
6	1309	41.9	315	23	ABE89881 Human polypeptide
7	1309	41.9	316	22	AAE74101 Human colon cancer
8	1304	41.7	290	21	AAE24470 Human secreted pro
9	1187.5	38.0	622	22	AAE79050 Human protein SEQ

10	1187.5	38.0	622	22	AAE38859 Human polypeptide
11	1187.5	38.0	632	22	AAE80034 Human protein SEQ
12	1187.5	38.0	632	22	AAE40645 Human polypeptide
13	1185.5	37.9	622	22	AAE93285 Human polypeptide,
14	1182.5	37.9	622	23	ABJ05568 Breast cancer-asso
15	1142.5	36.6	633	19	AAE34470 Human N-acetylgl
16	1122.5	35.9	556	23	ABG32379 Novel human enzyme
17	1101.5	35.3	561	21	AAE43561 Human cancer assoc
18	1100.5	35.2	559	16	AAE66397 Cattle GalNAC-tran
19	1100.5	35.2	559	16	AAE66401 GalNAC-transferase
20	1100.5	35.2	559	18	AAE16484 Bovine N-acetylgl
21	1100	35.2	517	16	AAE66402 GalNAC-transferase
22	1100	35.2	517	18	AAE16489 Honeybee mellitin
23	1076.5	34.5	630	22	ABE64271 Drosophila melanog
24	1058	33.9	478	24	ABE41518 Human DTHP protei
25	1057	33.8	590	22	ABE59763 Drosophila melanog
26	1023	32.7	639	21	AAE99440 Human PRO1564 (UNO
27	1023	32.7	639	22	AAE29205 Human PRO polypept
28	1023	32.7	639	22	AAE73178 UDP-GalNac: polype
29	1023	32.7	639	22	AAE66189 Protein of the inv
30	1023	32.7	639	23	ABE64882 Human protein SEQ
31	1023	32.7	639	23	ABE96444 Human ovarian can
32	1023	32.7	639	24	ABU71293 Human PRO1564 prot
33	1023	32.7	639	24	ABE65750 Human secreted/tra
34	1023	32.7	639	24	ABE66083 Novel human secret
35	1023	32.7	639	24	ABU67587 Human secreted/tra
36	1023	32.7	639	24	ABU65445 Human PRO polypept
37	1023	32.7	639	24	ABU58581 Human PRO polypept
38	1023	32.7	639	24	ABU56117 Human secreted/tra
39	1023	32.7	639	24	ABU57112 Human PRO polypept
40	1017	32.6	638	24	ABU10691 Human secreted/tra
41	1017	32.6	638	21	AAE78352 Human ZAP-3 protei
42	1013.5	32.4	571	22	AAE24054 Human EST encoded
43	1012	32.4	570	16	AAE83053 N-acetylglactosam
44	991.5	31.7	603	22	AAU07778 Human novel transf
45	991.5	31.7	603	23	AAU99098 Human glycosyl tra

ALIGNMENTS

RESULT 1
ABG32510
ID ABG32510 standard; Protein; 581 AA.

XX AC ABG32510;

XX DT 15-NOV-2002 (first entry)

XX XX Human glycosyltransferase 33945.

XX DE Human, enzyme; glycosyltransferase; 33945; atherosclerosis;
XX KW cardiovascular disorder; ischaemia; atherosclerosis; cancer; tumour;
XX KW congestive heart failure; endothelial cell disorder; psoriasis;
XX KW diabetic retinopathy; angina; hypertension; atrial fibrillation;
XX KW valvular disease; cardiomyopathy; haemangioma; pancreatic disorder;
XX KW cellular proliferative disorder; differentiative disorder; diabetes;
XX KW autoimmune disorder; haematopoietic neoplastic disorder; leukaemia;
XX KW Hodgkin's disease; chronic myelogenous leukaemia; inflammatory disease;
XX KW arthritis; multiple sclerosis; viral infection; liver disorder;
XX KW liver fibrosis; hepatocellular cancer.

XX OS Homo sapiens.

XX PN WO200264815-A2.

XX PD 22-AUG-2002.

XX PF 14-FEB-2002; 2002WO-US05042.

XX PR 15-FEB-2001; 2001US-269202P.

XX XX (MILL-) MILLENNIUM PHARM INC.

XX PA

XX Olandt PJ, Meyers RE, Galvin KA;
XX WPI; 2002-643491/69.
DR N-PSDB; ABS52469.
XX New 33945 nucleic acid molecule and polypeptide, useful for diagnosing
XX PT and treating disorders involving aberrant or deficient
XX glycosyltransferase function or expression such as atherosclerosis or
XX endothelial cell disorders -
XX
XX Claim 8; Page 100-101; 107pp; English.
XX
XX The invention relates to an isolated 33945 nucleic acid molecule encoding
XX a glycosyltransferase polypeptide, its fragment or allelic variant
XX and the encoded protein. Also included are host cells, antibodies,
XX producing the protein from cell culture, detecting the
XX protein/nucleic acid using probes or binding compounds and identifying
XX compounds (modulators) which bind the protein. The methods and
XX compositions of the present invention are useful for diagnosing and
XX treating disorders involving aberrant or deficient glycosyltransferase
XX function or expression such as atherosclerosis, cardiovascular
XX disorders (e.g. ischaemia, atherosclerosis, congestive heart failure)
XX endothelial cell disorders (e.g. tumours, psoriasis, diabetic
XX retinopathy) angina, hypertension, atrial fibrillation,
XX valvular disease, cardiomyopathy, haemangiomas, cancers, pancreatic
XX disorders, cellular proliferative and/or differentiative disorders,
XX autoimmune disorders, haematopoietic neoplastic disorders (e.g.
XX leukaemia, Hodgkin's disease, chronic myelogenous leukaemia),
XX inflammatory diseases (e.g. diabetes, arthritis, multiple sclerosis),
XX viral infection (e.g. Hepatitis B, hepatitis C and herpes simplex
XX virus) and liver disorders (e.g. liver fibrosis and hepatocellular
XX cancer). Many more diseases and disorders are listed in the
XX specification. The present sequence represents the glycosyltransferase
XX 33945.
XX
XX Sequence 581 AA;
XX
XX Query Match 100.0%; Score 3124; DB 23; Length 581;
XX Best Local Similarity 100.0%; Pred. No. 8.6e-292;
XX Matches 581; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MWGTTARRRRPRELRRGREALLVLLALLAGLGSVLRQAAGAGAAEPGPPRTPGR 60
XX
XX 1 MWGTTARRRRPRELRRGREALLVLLALLAGLGSVLRQAAGAGAAEPGPPRTPGR 60
XX
XX 61 REPVMPRPVPANALGARGEAVRLQGEELRLOEESVRLHQINILYSDRISLHRLPER 120
XX
XX 61 REPVMPRPVPANALGARGEAVRLQGEELRLOEESVRLHQINILYSDRISLHRLPER 120
XX
XX 121 WNPLCKEKKYDYNLPRTSVITIAFYNEAWSTLLRTVYVLETSDDILLEEVILVDDYSDR 180
XX
XX 121 WNPLCKEKKYDYNLPRTSVITIAFYNEAWSTLLRTVYVLETSDDILLEEVILVDDYSDR 180
XX
XX 181 EHLKERLANELSGPKVRLIRANKREGVLRARLGCASARGDVLTFDLCHEGWLPEP 240
XX
XX 181 EHLKERLANELSGPKVRLIRANKREGVLRARLGCASARGDVLTFDLCHEGWLPEP 240
XX
XX 241 LLQRIHEESAVVCPVIDVDMNTFEYLGNSGEPQIGFDWRLVFTWHTPERERIRMQS 300
XX
XX 241 LLQRIHEESAVVCPVIDVDMNTFEYLGNSGEPQIGFDWRLVFTWHTPERERIRMQS 300
XX
XX 301 PVDVIRSPHTAGGLFAVSKKYFEYLGSDYDTGMEVWGGENLEFSFRWQCGGVLETHPCSH 360
XX
XX 301 PVDVIRSPHTAGGLFAVSKKYFEYLGSDYDTGMEVWGGENLEFSFRWQCGGVLETHPCSH 360
XX
XX 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEKELYIYHNPRALEPFGDVTTERKQLRDKL 420
XX
XX 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEKELYIYHNPRALEPFGDVTTERKQLRDKL 420
XX
XX 421 QCKDFKWFLETVPYELHVPEDRPGFFGMQLNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
XX
XX 421 QCKDFKWFLETVPYELHVPEDRPGFFGMQLNKGLTDYCFDYNPPDENQIVGHQVILYLCH 480
XX

QY 481 GMGQNOFFEYTSQKEIRYNTHQPEGCI AVEAGMDTLMHLCEETAPENQKFIQEDGSLF 540
DB 481 GMGQNOFFEYTSQKEIRYNTHQPEGCI AVEAGMDTLMHLCEETAPENQKFIQEDGSLF 540
QY 541 HESQKVCQAARKESSDSFVPLLRDCTNSDQKQWFFKERM 581
DB 541 HESQKVCQAARKESSDSFVPLLRDCTNSDQKQWFFKERM 581

RESULT 2
AAE25019
ID AAE25019 standard; Protein; 581 AA.
XX AC AAE25019;
XX DT 30-OCT-2002 (first entry)
XX DE Human drug metabolising enzyme (DME-4).
XX KW Human; drug metabolising enzyme; autoimmune; inflammatory disorder;
XX KW acquired immunodeficiency syndrome; AIDS; atherosclerosis; psoriasis;
XX KW proliferative disorder; arteriosclerosis; cirrhosis; hepatitis; cancer;
XX KW dementia; neurological disorder; Alzheimer's disease; Huntington's disease;
XX KW adenoma; Parkinson's disease; developmental disorder; anaemia; adenoma;
XX KW drug screening; endocrine disorder; conjunctivitis; glaucoma; cataract;
XX KW renal tubular acidosis; eye disorder; epilepsy; thrombosis; peptic ulcer;
XX KW anorexia; metabolic disorder; cystic fibrosis; diabetes; liver disorder;
XX KW goitre; gastrointestinal disorder; gene therapy; virucide; anticoagulant;
XX KW anticonvulsant; nootropic; enzyme; DME-4.
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX Peptide 1..34
XX Protein /label= Signal_peptide
XX Peptide /note= "Mature human DME-4"
XX Protein 35..581
XX Peptide 1..39
XX Protein /label= Signal_peptide
XX Peptide 40..581
XX Protein /note= "Mature human DME-4"
XX Domain 14..41
XX /note= "Transmembrane domain"
XX WO200246426-A2.
XX
XX 13-JUN-2002.
XX
XX 04-DEC-2001; 2001WO-US47429.
XX
XX 08-DEC-2000; 2000US-254308P.
XX 15-DEC-2000; 2000US-256189P.
XX 21-DEC-2000; 2000US-257713P.
XX 19-JAN-2001; 2001US-262706P.
XX 02-FEB-2001; 2001US-266020P.
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Sanjanwala MM, Yao MG, Au-Young J, Baughn MR, Arvizu C, Ring HZ;
XX Lee EA, Ding L, Hafalia AJA, Tang YT, Yue H, Tribouley CM;
XX Lu DAM, Lal PG, Warren BA, Yang J, Wallia NK, Nguyen DB;
XX Gandhi AR, Lu Y, Ison CH;
XX WPI; 2002-519668/55.
XX N-PSDB; AAD40568.
XX
XX Novel human drug metabolizing polypeptide, useful in diagnosis,
XX prevention or treatment of autoimmune/inflammatory, cell proliferative,
XX neurological, developmental, endocrine, metabolic and gastrointestinal
XX disorders -
XX Claim 59; Page 146-147; 169pp; English.
XX PS

XX The invention relates to an isolated human drug metabolising enzyme (DME)
CC and its nucleotide. DME is useful for diagnosing, treating or preventing
CC disorders associated with aberrant expression of DME, where the disorders
CC are selected from autoimmune/inflammatory disorder such as acquired
CC immunodeficiency syndrome (AIDS), asthma, atherosclerosis, psoriasis,
CC uveitis; a cell proliferative disorder such as arteriosclerosis,
CC cirrhosis, hepatitis, and cancer; a neurological disorder such as
CC Alzheimer's disease, Huntington's disease, dementia, Parkinson's disease;
CC a developmental disorder such as renal tubular acidosis, epilepsy,
CC anaemia; an endocrine disorder such as adenoma, thrombosis and
CC infections; an eye disorder such as conjunctivitis, glaucoma, cataract;
CC metabolic disorder such as cystic fibrosis, diabetes and goitre; a
CC gastrointestinal disorder such as anorexia, peptic ulcer; and liver
CC disorders. DME is useful in a number of drug screening techniques and to
CC analyse the proteome of a tissue or cell type. The invention is useful
CC for creating knock-in humanised animals or transgenic animals to model
CC human diseases, in somatic or germline gene therapy, to generate a
CC transcript image of a tissue or cell type, for detecting differences in
CC the chromosomal location due to translocation, inversion, etc. among
CC normal, carrier or affected individuals, and as hybridisation probes for
CC mapping naturally occurring genomic sequences. The present sequence is
CC human DME-4.
XX
SQ Sequence 581 AA;

Query Match 100.0%; Score 3124; DB 23; Length 581;
Best Local Similarity 100.0%; Pred. No. 8.6e-292;
Matches 581; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MWGTARRCPRELRGREALLVLLALLAGLSVLAQRGAGAAEPGPPRPRGR 60
Db 1 MWGTARRCPRELRGREALLVLLALLAGLSVLAQRGAGAAEPGPPRPRGR 60
Qy 61 REPVMPPPANALGARGEAVRLQGEELRLOEESVRLHQINILSDRSLHRRPER 120
Db 61 REPVMPPPANALGARGEAVRLQGEELRLOEESVRLHQINILSDRSLHRRPER 120
Qy 121 WNPCKEKKDYDNLPTSIIAFYNEAWSLTLLRTVSVLETSPIILLEEVILVDDYSDR 180
Db 121 WNPCKEKKDYDNLPTSIIAFYNEAWSLTLLRTVSVLETSPIILLEEVILVDDYSDR 180
Qy 181 EHLKERLANELSGLPKVLIRANKREGILVRALLGASAAAGDVLTFDCHCEHGWLEP 240
Db 181 EHLKERLANELSGLPKVLIRANKREGILVRALLGASAAAGDVLTFDCHCEHGWLEP 240
Qy 241 LLQRIHEESA VCPVIDVDWNTPEYLGNSGEPOIGGFDWRLVFTWHTVPERIRMQS 300
Db 241 LLQRIHEESA VCPVIDVDWNTPEYLGNSGEPOIGGFDWRLVFTWHTVPERIRMQS 300
Qy 301 PVDVIRPTWAGGLFAVSKKYFEYLGSDYDTGMEVWGGENLEFSPRIWCCGVLETHPCSH 360
Db 301 PVDVIRPTWAGGLFAVSKKYFEYLGSDYDTGMEVWGGENLEFSPRIWCCGVLETHPCSH 360
Qy 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEKELYHYHNPRLARLEPGDVTERRKQDRKL 420
Db 361 VGHVFPKQAPYSRNKALANSVRAAEVWMDPEKELYHYHNPRLARLEPGDVTERRKQDRKL 420
Qy 421 QCKDFKWFLETVYELHVPEDRPGFGLQNGKLTDCFDYNDPNDENQIVGHQVILYLCH 480
Db 421 QCKDFKWFLETVYELHVPEDRPGFGLQNGKLTDCFDYNDPNDENQIVGHQVILYLCH 480
Qy 481 GMGNQPFYTSQKEIRYNTHQPGGCI AVEAGMDTLIMHLCETAPENQKFIQEDGSLF 540
Db 481 GMGNQPFYTSQKEIRYNTHQPGGCI AVEAGMDTLIMHLCETAPENQKFIQEDGSLF 540
Qy 541 HEQSKVCQAARKESSDSFVPLLRDCTNSDHQKWFKERM 581
Db 541 HEQSKVCQAARKESSDSFVPLLRDCTNSDHQKWFKERM 581

RESULT 3
ABG04767

ID ABG04767 standard; Protein; 355 AA.
XX
AC ABG04767;
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #4758.
XX
DE Human chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR
DR N-PSDB; AAS68954.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 35126; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG0377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 355 AA;
Query Match 57.2%; Score 1788; DB 22; Length 355;
Best Local Similarity 98.5%; Pred. No. 2.5e-163;
Matches 335; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 65 MPRPPVPANALGARGEAVRLQGEELRLOEESVRLHQINILSDRSLHRRPERWNPL 124
Db 1 MPRPPVPANALGARGEAVRLQGEELRLOEESVRLHQINILSDRSLHRRPERWNPL 60
Qy 125 CKEKKYDYNLPTSIIAFYNEAWSLTLLRTVSVLETSPIILLEEVILVDDYSDRHLK 184
Db 61 CKEKKYDYNLPTSIIAFYNEAWSLTLLRTVSVLETSPIILLEEVILVDDYSDRHLK 120
Qy 185 ERLANELSGLPKVLIRANKREGILVRALLGASAAAGDVLTFDCHCEHGWLEPLQR 244

Db 121 ERLANELSGLPKVKRLIRANKREGLVRARLLGASAAAGDVLTFDLDCHCEHGLEPQLQR 180
 Qy 245 IHEESAVVCVIDVIDWNTFEYILGNSGEPOIGGFDHRLVFTWHTVPERERIRMQSPVDV 304
 Db 181 IHEESAVVCVIDVIDWNTFEYILGNSGEPOIGGFDHRLVFRWHTVPERERIRMQSPVDV 240
 Qy 305 IRSPTMAGGLFAVSKKYFEYILGSDYDTGMEVWGGNLEFSFRIMWCGGVLETHPCSHVGHV 364
 Db 241 IRSPTMAGGLFAVSKKYFEYILGSDYDTGMEVWGGNLEFSFRIMWCGGVLETHPCSHVGHV 300
 Qy 365 FPKQAPYSRNKALANSVRAAEVWMDPEFKELYHHRNPRARL 404
 Db 301 FPKQAPYSRNKALANSVRAAEVWMDPEFKELYHHRNPRARL 340

RESULT 4

ABU07527
 ID ABU07527 standard; Protein; 578 AA.
 AC ABU07527;
 XX
 DT 13-MAR-2003 (first entry)
 XX
 DE Human N-acetylgalactosaminyl transferase T4, GalNac T4.
 XX
 KW Human; enzyme; GalNac T4; N-acetylgalactosaminyl transferase T4;
 KW glycosylation; mucin 1; MUC1; vaccine; antiinflammatory; GalNac-T1;
 KW GalNac-T2; GalNac-T3.
 XX
 OS Homo sapiens.
 XX
 PN US646S220-B1.
 XX
 PD 15-OCT-2002.
 XX
 PF 21-DEC-1998; 98US-0217306.
 XX
 PR 21-DEC-1998; 98US-0217306.
 XX
 PA (GLYC-) GLYCOZYM APS.

XX Hassan FH, Clausen H, Bennett EP, Eisenkraetzer D, Gaetgens J;
 PI WPI; 2003-147066/14.
 DR N-PSDB; ABX14962.
 XX

PT Glycosylating MUC1 acceptor substrate, by glycosylating substrate with
 PT N-acetylgalactosaminyltransferase T1, GalNac-T2 or GalNac-T3, then with
 PT human GalNac-T4 to glycosylate specific Ser, Thr residues in substrate
 PT

XX Claim 2; Column 11-16; 10pp; English.

XX The invention relates to glycosylating a MUC1 (mucin 1) acceptor
 CC substrate, comprising glycosylating the substrate with enzymatically
 CC active N-acetylgalactosaminyltransferase (GalNac)-T1, GalNac-T2 or
 CC GalNac-T3, or with GalNac capable of glycosylating MUC1 glycosylation
 CC sites that can be glycosylated by GalNac-T1, GalNac-T2 or GalNac-T3, and
 CC glycosylating the substrate with enzymatically active human GalNac-T4 to
 CC glycosylate specific Ser, Thr positions in the MUC1 substrate.
 CC The method is used for glycosylating an MUC1 acceptor substrate. The
 CC glycosylated substrates are useful in preparation of vaccines and
 CC antiinflammatory agents. GalNac-T4 exhibits a different substrate
 CC specificity than previously characterised GalNac transferases. The
 CC activity of GalNac-T4 is unique and specific to glycosylate specific
 CC serine and threonine residues in MUC1 tandem repeat.
 CC The present sequence is the human GalNac T4 protein.

XX Sequence 578 AA;

Query Match 52.4%; Score 1638; DB 24; Length 578;
 Best Local Similarity 56.8%; Pred. No. 1.5e-148;

Matches 324; Conservative 78; Mismatches 150; Indels 18; Gaps 8;
 Qy 23 VLLALLALAG-----LGSVLRAQRGAGAGAAEPPTPRGRR-----EPVMPRPDPVA 72
 Db 13 LLLAFITVAYIFVELLVSTFASAGAGR-ARELGSRRSLDLQKNTEDLSRLYKPKPADS 71
 Qy 73 NALGARGEAVRLQLOGEELRLQEEVRLHQINTYLSDRISLHRLRLPERMNPCKEKKYDY 132
 Db 72 RALGEGKASKLQNLDELKQEEELIETVAINIYLSDRISLHRIEDKRWYCKESQKFN 131
 Qy 133 DNPRTSVIIAFNEAWSTLLRTVYSVLETSPIILLSEVILVDDYSDREHLKERLANELS 192
 Db 132 RTLEPTTSVITAFNEAWSTLLRTIHSVLETSPIALLKEIILVDDLSDRVLYLKTQLEYIS 191
 Qy 193 GLPKVRLIRANKREGLVRARLLGASAAAGDVLTFDLDCHCEHGLEPQLQR IHEESAV 252
 Db 192 NLDVRVRLIRNKREGLVRARLLGATFATGDLVTLFYCHCECNSGWLPLELLERIGRYTAV 251
 Qy 253 VCPVIDVDWNTFEYILGNSGEPOIGGFDHRLVFTWHTVPERERIRMQSPVDVIRSPTMAG 312
 Db 252 VCPVIDTIDWNTFEYIMQIGEPWIGGFDHRLVFTQWHSVPKQERDRIRISIDPISPTMAG 311
 Qy 313 GLFAVSKKYFEYILGSDYDTGMEVWGGNLEFSFRIMWCGGVLETHPCSHVGHVFPKQAPYS 372
 Db 312 GLFAVSKKYFPYLGTYDTGMEVWGGNLELSFRWQCGGKLETHPCSHVGHVFPKRAPYA 371
 Qy 373 RNKALANSVRAAEVWMDPEFKELYHHRNPRARLPFGDVTERRKOLRDLCKDFKWFLETV 432
 Db 372 RPNFLQNTARAAEVWMDPEYKEHFYNNPPARKAYGDISERKLLRELRCKSPDWYLNKV 431
 Qy 433 YPELHVPEDRPGFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCHGNGQNGOFFEYTS 492
 Db 432 FPNLHVPEDRPGHGAIRSGISSECLDYNSPDNNP-TGANLSLFGCHGCGGNGOFFEYTS 490
 Qy 493 QKEIRYNTHQPEGCIAVEAGMDLIMHLCEE---TAPENQKFILOEPDGLSFLHEQSKCVQ 549
 Db 491 NKEIRFNS-VTELCAEVPEQNYVMQNCPKDGPVPANIWHFKEDGTIFPHSHGLCLS 549
 Qy 550 AARKESSDSFVPLLRDCTNSD-HQKWFPEKE 578
 Db 550 AYTPGGRPDVQ-MRTCDALDKNQIWSFEK 578

RESULT 5

AAB53452
 ID AAB53452 standard; Protein; 256 AA.
 XX

XX AAB53452;

XX AC
 XX DT 09-MAR-2001 (first entry)
 XX

XX Human colon cancer antigen protein sequence SEQ ID NO:992.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder.

XX Homo sapiens.

XX WO200005351-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05883.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX


```
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
XX
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI; 2001-235357/24.
XX
XX N-PSDB; AAH33532.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 11; Page 6646-6648; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing P.
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAH77789 represent sequences used in the exemplification of the
XX present invention.
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 316 AA;
XX
XX Query Match 41.9%; Score 1309; DB 22; Length 316;
XX Best Local Similarity 99.6%; Pred. No. 3.2e-117;
XX Matches 236; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 345 RIWQCGVLETHPCSHVGHVFPKQAPYSRNKALANSVRAAEVWMDPEKELYHNRPRARL 404
XX :|||||
XX 80 QIWQCGVLETHPCSHVGHVFPKQAPYSRNKALANSVRAAEVWMDPEKELYHNRPRARL 139
XX :|||||
XX 405 EPGDVTERKQLRDKLQCKDFKWFLETVYPHLPVDPDRPGFGMLQNKGLTDYCFDYNPP 464
XX :|||||
XX 140 EPGDVTERKQLRDKLQCKDFKWFLETVYPHLPVDPDRPGFGMLQNKGLTDYCFDYNPP 199
XX :|||||
XX 465 DENQIVGHQVILYLCHGNGQNFPEYTSQKEIRYNTHQPEGCIAVEAGMDTLIMHLCSET 524
XX :|||||
XX 200 DENQIVGHQVILYLCHGNGQNFPEYTSQKEIRYNTHQPEGCIAVEAGMDTLIMHLCSET 259
XX :|||||
XX 525 APENQKFILODGSILFHEQSKKCVQAAKESDSDVPLLRDCTNSDHQKWFKERMML 581
XX :|||||
XX 260 APENQKFILODGSILFHEQSKKCVQAAKESDSDVPLLRDCTNSDHQKWFKERMML 316
XX :|||||
XX
XX RESULT 8
XX AAB24470
XX ID AAB24470 standard; Protein; 290 AA.
```

```
XX AAB24470;
XX
XX 20-NOV-2000 (first entry)
XX
XX Human secreted protein sequence encoded by gene 34 SEQ ID NO:95.
XX
XX Human; secreted protein; cytostatic; antianaemic; antidiabetic;
XX antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
XX antipsoriatic; antiangiogenic; cardiant; anti-HIV; nootropic;
XX neuroprotective; antimicrobial; antiparkinsonian; cancer;
XX immune system disorder; angiogenesis; hyperproliferative disorder;
XX cardiovascular disorder; apoptosis; neurologic disease;
XX infectious disease; wound healing.
XX
XX Homo sapiens.
XX
XX WO200035937-A1.
XX
XX 22-JUN-2000.
XX
XX 16-DEC-1999; 99WO-US29950.
XX
XX 17-DEC-1998; 98US-0112809.
XX
XX 18-DEC-1998; 98US-0113006.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Ebner R, Rosen CA, Endress GA, Soppet DR, Ni J;
XX Duan DR, Moore PA, Shi Y, Lafleur DW, Olsen HS, Florence K;
XX
XX WPI; 2000-431566/37.
XX
XX N-PSDB; AAH78414.
XX
XX Forty seven human nucleic acids encoding secreted proteins, useful in
XX the treatment, prevention and diagnosis of cancers, disorders of the
XX immune system, angiogenesis disorders, neurological diseases and
XX hyperproliferative disorders -
XX
XX Claim 11; Page 514-515; 562pp; English.
XX
XX The polynucleotide sequence given in AAA78381 to AAA78432 encode the
XX human secreted proteins given in AAB24437 to AAB24604. Human secreted
XX proteins have activities based on the tissues and cells the genes are
XX expressed in. Examples of activities include: cytostatic; antianaemic;
XX antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
XX antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
XX neuroprotective; antimicrobial and antiparkinsonian.
XX Human secreted protein polynucleotides, polypeptides, antagonists and/or
XX agonists may be useful in treating, preventing, and/or diagnosing other
XX diseases, disorders, and/or conditions such as: (a) cancers; (b)
XX disorders of the immune system; (c) angiogenesis disorders; (d)
XX hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
XX associated with increase apoptosis; (g) neurological diseases; and
XX (h) infectious diseases. They are also used to promote wound healing.
XX AAA78372 to AAA78380 and AAB24436 represent sequences used in the
XX exemplification of the present invention.
XX
XX Sequence 290 AA;
XX
XX Query Match 41.7%; Score 1304; DB 21; Length 290;
XX Best Local Similarity 99.2%; Pred. No. 8.5e-117;
XX Matches 235; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX 345 RIWQCGVLETHPCSHVGHVFPKQAPYSRNKALANSVRAAEVWMDPEKELYHNRPRARL 404
XX :|||||
XX 53 QIWQCGVLETHPCSHVGHVFPKQAPYSRNKALANSVRAAEVWMDPEKELYHNRPRARL 112
XX :|||||
XX 405 EPGDVTERKQLRDKLQCKDFKWFLETVYPHLPVDPDRPGFGMLQNKGLTDYCFDYNPP 464
XX :|||||
XX 113 EPGDVTERKQLRDKLQCKDFKWFLETVYPHLPVDPDRPGFGMLQNKGLTDYCFDYNPP 172
XX :|||||
XX 465 DENQIVGHQVILYLCHGNGQNFPEYTSQKEIRYNTHQPEGCIAVEAGMDTLIMHLCSET 524
XX :|||||
XX
```



```
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI58015.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 3; SEQ ID NO 2004; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 622 AA;

Query Match 38.0%; Score 1187.5; DB 22; Length 622;
Best Local Similarity 45.9%; Pred. No. 4.6e-165;
Matches 249; Conservative 91; Mismatches 103; Indels 39; Gaps 14;

QY 58 PGRREPVPVPANALGARGEAVR---LQLQGEELRLQESVRLHQINIVLSRISL 113
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 96 PAELKPFWERPPQPNAGADGKAFQSKWTPLTQEQ---KEEGYKXKCFNAPASDRISL 152
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 114 HRLR-PRRNPLCKEKY-DVDNLPRTSVIIAFVNEAWSLLRTTVYSVLETSPPDILEEV 171
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 153 QRSGLPDRPPCEVDQKFRRCPLATTSVIIVFHNEAWSLLRTTVYSVLETSPPAILKEI 212
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 172 ILVDYSDREHLKERLANELSLKPLRLIRANKREGVLRAELGASARGDVLTLFDDHC 231
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 213 ILVDASTEEHLKEKLEQYVKQLQVRVVRQEEERKGLITARLLGASVAQAELVTLFDDHC 272
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 232 ECHGWLPLELQRIHEESAVVCPVIDVDWNTPEYLGNSGEPI---GGFDMRLVFTWH 288
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 273 ECFHGLEPLLARAEDKTVVSDVIDTIDNTEFAKPVGRVHSGNFDLSLTGWE 332
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 289 TVPERIRMQSPVDVIRSPMTMAGLFAVSKYFYLGSYDTGMVNGENLEFSFRWQ 348
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 333 TLPPEKQRRKDETVPIKSPTFAGLFSISKSYFEHIGTYNQMEIWMGGENVMSFRWQ 392
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 349 CGGVLTHPCSHVGHVFPKQAPYSRNLKALA---NSVRAAEVWVDEKELYHFN---P 400
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 393 CGGQLEIIPCWSVGHVFPKQAPYSRNLKALA---NSVRAAEVWVDEKELYHFN---P 452
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 401 RARLEPFGDVTERKQRLDKCKPFWFETVYVPELHVPEDRPGFFGMQLNGKGLTDYCFD 460
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 453 MAQESKSGDIERLQRLQCHNPSWYLNHNVPEMFVPLDTPFYGAIKNLG-TNQCLD 511
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 461 YNPDPENQVGHVTLVYCHGNGQNPFEYTSQKEIRVYNTHOPEGCTIAVEAGMTLIMHL 520
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 512 VG---ENNRGGKPLIMYSCHGLGGNQYFEYTTQDRLRHIAK-QLCLHVSXG--ALGLGS 565
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

521 CBET-----APENQKFILOEDGSLFHEQSKKCVQARKESSDSFVFLLRDCTNSD-HQKW 574
566 CHFTGKNSQVPKDEEWELAQDLIRNSGSGTCTUTSQDKK-----PAMAPCNPSDPQLW 619
575 FF 576
620 LF 621

RESULT 11
AAM80034
ID AAM80034 standard; Protein; 632 AA.
XX
AC AAM80034;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3680.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WQ200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
(HYSE-) HYSEQ INC.
XX
Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX
WPI; 2001-476283/51.
DR N-PSDB; AAK53167.
XX
Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
Claim 20; Page 411-412; 6221pp; English.
XX
The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 632 AA;

Query Match 38.0%; Score 1187.5; DB 22; Length 632;
```

```
Best Local Similarity 45.9%; Pred. No. 4.7e-105;
Matches 249; Conservative 91; Mismatches 163; Indels 39; Gaps 14;

Qy 58 PGRREPMPRPVPANALGARGEAVR---LQLOGEELRLQEEVSRVLRHQINILYLSDRISL 113
Db 106 PAELKPFWRPPQDPNAPGADGKAFQSKWTPLETQE---KEEGYKXCFNAPASDRISL 162

Qy 114 HRRL-PRWNPCKEKKY-DYDNLPRTSVIAFYNEAWSMTLLRTVYSLTSPDILLEEY 171
Db 163 QRSLGPDTRPPECVDQKFRCPPLATTSVIVFHNEAWSMTLLRTVYSLTTPAILLKEI 222

Qy 172 ILVDYSDREHLKERLANELSGLPKVRIRANKREGLVRARLIGASAAAGDVLTLFDC 231
Db 223 ILVDASTEEHLKEKLEQYVQKQVVRVVRQERKGLITARLLGASVAQAEVLTLDAHC 282

Qy 232 ECHGWLPLELQRIHEESAVVCPVIDVIDWNTFEYLGNSGEPQI---GGFDMRLVFTWH 288
Db 283 ECFHGMLEPLARIAEDKTVVSPDIVTIDLNTFEPAKPVQGRVHSGRGNFDSLTGWE 342

Qy 289 TVPRERIRMQSPVDVIRSPMTAGGLFAVSKKYFEYLGSDYTGMEVWGENLEFSFRWQ 348
Db 343 TLPPEKQRRKDETYPIKSPTFAGGLFSISKSYFEHIGTYDQMEIWGENVENFSFRWQ 402

Qy 349 CGGVLETHPCSHVGHVPKQAPYSRNKALA---NSVRAAEVWVDEFEKELYHNRN---P 400
Db 403 CGGOLEIIPCSVGHVFRTPKSPHTFPKGTSVIARNQVRLAEVWVDSYKXIFYRNLOAAK 462

Qy 401 RARLEPFGDVTERRKQRLDKQCKDFKWFLETVYPPELVHVPDRPGFFGMLQNGKLTDCFD 460
Db 463 MAQEKSGDISERLQRLREQLHCHNFNSWYLVNVPFVDPDLTPFYGAIKNLG-TNQCILD 521

Qy 461 YNPDPENQIVGHQVILYLCHEGMQNGQFFEYTSQKEIRYNTHQPEGCIATVAGMDTLIMHL 520
Db 522 VG---ENNRGGKPLIMYSCHGLGNQYFEYTTQDRLRHNTAK-QLCLHVSXG--ALGLGS 575

Qy 521 CEET-----APENQKFILOEDGSLFHEQSKKCVQAAKESDVSFVPLLRDCTNSD-HQKW 574
Db 576 CHFTGKNSQVPKDBEWELAOPLIRNSGSGTCLTSQDKK-----PAMAPCNPSDPHQLW 629

Qy 575 FF 576
Db 630 LF 631

RESULT 12
AAM40645
ID AAM40645 standard; Protein; 632 AA.
AC AAM40645;
XX
XX
XX 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 5576.
XX
XX Human; nototropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX Homo sapiens.
XX
XX W0200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX
XX 25-APR-2000; 2000US-0552317.
XX
XX 09-JUL-2000; 2000US-0598042.
XX
XX 19-JUL-2000; 2000US-0620312.
```

```
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
XX N-PSDB; AAI59801.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 5576; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nototropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX SQ Sequence 632 AA;
XX
XX Query Match 38.0%; Score 1187.5; DB 22; Length 632;
XX Best Local Similarity 45.9%; Pred. No. 4.7e-105;
XX Matches 249; Conservative 91; Mismatches 163; Indels 39; Gaps 14;
```

QY 521 CEET-----APENOKFILOEDGSLFHEQSKCVQAAKESSDSFVPLLRDCTNSD-HQKW 574
DB 576 CFTGKNSQVPKDEWELAQDLIRNSGSGTCLISQDKK-----PAMAPCNPSDPHQLW 629

QY 575 FF 576
DB 630 LF 631

RESULT 13
AAM93285
ID AAM93285 standard; Protein; 622 AA.
XX AAM93285;
AC AAM93285;
DT 06-NOV-2001 (first entry)
XX Human polypeptide, SEQ ID NO: 2769.
XX Human; full length cDNA; cDNA synthesis; oligo-capping.
XX Homo sapiens.
XX EP1130094-A2.
XX 05-SEP-2001.
XX 07-JUL-2000; 2000EP-0114089.
XX 08-JUL-1999; 99JP-0194486.
XX 11-JAN-2000; 2000JP-0118774.
XX 02-MAY-2000; 2000JP-0183765.
XX (HELI-) HELIX RES INST.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
XX N-PSDB; AAK94205.
XX 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -
XX Claim 8; SEQ ID NO 2769; 1380pp + sequence listing; English.
XX The invention relates to primers for synthesising full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesising the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is a polypeptide
XX encoded by a full length human cDNA of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.
XX SQ Sequence 622 AA;

Query Match 37.9%; Score 1185.5; DB 22; Length 622;
Best Local Similarity 45.9%; Fred. NO. 7.1e-105;
Matches 249; Conservative 91; Mismatches 163; Indels 39; Gaps 14;

QY 58 GRRREPVRPPVPANALGARGEAVR---LQOGEELRLOESVRLHGINIYLSDRISL 113
DB 96 PAELKPFWRPPQDNACGAKGAFQKSKWTPLETQE---KEEGYKCKCFNAPASDRISL 152

QY 114 HRLR-PERNPLCKEKKY-DYDNLPRTSVIAFYNEAWSTLLRTVYSVLETSPIILUEV 171
DB 153 ORSLGPDTRPPECVDQKFRCPPLATTSSVIVFHNEAWSTLLRTVYSVLTHTTPIILKEI 212

QY 172 ILVDDYSDBRHLKERLANELSGLPKVLIRANKREGVLRRARLLGASAAARGDVLTFDLCHC 231
DB 213 ILVDDASTEHLEKELEQYVKQLQVVRVVRQBERKGLITARLLGASVAAAEVLTFDLAHC 272

QY 232 ECHGWLPLELLQRIHEESAVCPVIDIDWNTFEYLGNSGEPOI---GGFDWRLVFTWH 288
DB 273 ECFHGWLEPLLARIAEDKTWVSPDIVTIDLNTFEFAKPVQGRVHSRGNFDSLTCWE 332

QY 289 TVPERERIRMQSPVDVIRSPTMAGGLPAVSKKFEYLGSDYDTGMEVWGENLEPFRIMQ 348
DB 333 TLPPEHKQRKRDETPIKSPFAGGLFSISKSYFHEHGTIDYDQMEIWGENVENMSFRWQ 392

QY 349 CGGVLETHPCSHVGHVFPKQAPYSRNKALA---NSVRAAEVWMDPEKELYHHRNPA-- 402
DB 393 CGGLEIIPCSVGVGHVPRTPKSPHFPKGTSVIARNQVRLAEVWMDSKYFYRNLQAAK 452

QY 403 --RLEPFGDVTERKQLRDKLQCKFKWLFETVYVPELVHPEDRPGFFQMLQKNGLDYCFD 460
DB 453 MTQEKSGDISERLQRLREQLHCHNFSWYLNHVYPMFVPLTPTFYGAIKNLG-TNQLD 511

QY 461 YNPEDENQIVGHQVILYLCHGMGQNFPEYTSQKEIRYNTHOPEGCIATVAGMDTLIMHL 520
DB 512 VG---ENNRGGKPLIMYSCHGLGNGQYFEYTTQDLRHNTAK-RLCLHVSKG--ALGLGS 565

QY 521 CEET-----APENOKFILOEDGSLFHEQSKCVQAAKESSDSFVPLLRDCTNSD-HQKW 574
DB 566 CFTGKNSQVPKDEWELAQDLIRNSGSGTCLISQDKK-----PAMAPCNPSDPHQLW 619

575 FF 576
620 LF 621

RESULT 14
ABJ05568
ID ABJ05568 standard; Protein; 622 AA.
XX AC ABJ05568;
XX 14-NOV-2002 (first entry)
XX Breast cancer-associated protein 33.
XX Breast cancer; breast cancer-associated gene sequence;
XX drug development; pharmacogenetics; biosensor development.
XX Unidentified.
XX WO200259377-A2.
XX 01-AUG-2002.
XX 24-JAN-2002; 2002WO-US02242.
XX 24-JAN-2001; 2001US-263965P.
XX 02-FEB-2001; 2001US-265928P.
XX 09-APR-2001; 2001US-282698P.
XX 09-APR-2001; 2001US-0829472.
XX 04-MAY-2001; 2001US-288590P.
XX 29-MAY-2001; 2001US-294443P.
XX (EOSB-) EOS BIOTECHNOLOGY INC.
XX Mack DH, Gish KC, Afar D;
XX N-PSDB; ABT07725.
XX WPI; 2002-583738/62.
XX DR N-PSDB; ABT07725.
XX PT Detecting a breast cancer-associated transcript in a patient's cell,
XX useful for diagnosing breast cancer, comprises contacting a biological
XX sample with a polynucleotide that selectively hybridizes with breast
XX cancer nucleic acids -

Db	108	KPVLDRPPQDSNAPGASGKAFKTTNLSVEBQKEKGEAKHCFNAPASDRISLHRDLGPD	167
Qy	120	RWNPCKEKKYD-YDNLPTSVIIAFYNEAWSTLLRTVSVLSTSPDILLEEVLVDDYS	178
Db	168	TRPEECIEQFKRCPPLPPTTSVIIVFHNEAWSTLLRTVHVSYPAILLKEILLVDDAS	227
Qy	179	DREHLKERLANELSGLPKVPFLIRANKREGIARARLLGASARGDVLTFELDCHCEHEGWL	238
Db	228	VDEYLHDKLDEYVKQFSIVKIVRQRERKGLITARLLGATVATAETLTFELDAHCECFYGWL	287
Qy	239	EPLLORIHESAVVCVIDIVIDWNTFEYLGNSGEP-----QICGFDWRLVFTWHTVPER	293
Db	288	EPLLARIAENYTAUVSPDIASIDLNTFEF--NKPSPYGSHNHRGNFDWSLSFCWESLPDH	345
Qy	294	ERIRMQSPVDIRSPTNWAGLFAVSKKYFEYLSYDTGMVWGGENLEFSFRIWQCGVL	353
Db	346	EKORRKDETYPIKTPTTPTAGGLFSISKEYFEYIGSYDEEMEIWGGENIEMSFVWQCGQL	405
Qy	354	ETHPCSHVGHVFPKQAPYSRKA-----LANSVRAAEVWMDFEKELYVHRNPRA-----RLE	405
Db	406	EIMPCSVVGHVFRSKSPHSFPGTQVIARNQVRLAEVWMDXEYKEIFYRRNTDAKIVKQK	465
Qy	406	PFGDVTERRKOLRDKLCKDFKWFLETVYPSELHVPEDRPGFFGMLQNKGLTDYCFDYNPPD	465
Db	466	AFGLSKRFEIKHRLRCNFTWYLNNIYPEVYVFDLNPVISGYIKSVG-QPLCLDVG---521	
Qy	466	ENQIVGHQVILYLCHGMGQNPFEYTSQKEIRYN-----THOPEGCIAVEA-----GMDT	515
Db	522	ENNOGKPLIMYTCHGIGGNQYFEYSAQHEIRHNIQKELCLHAAQGLVQLKACTYKGHKT	581
Qy	516	LIMHLCEETAPENQKFLQEDGSLFHEQSKCVQAKESDSEFVPLLRDCTNSDH-QKW	574
Db	582	VW-----TGEIQIWEIQDKQLLYNPFLKXVCLSANGEH-----PSLVSCNPSDPLQKW	627
Qy	575	FFKE	578
Db	628	ILSQ	631


```

FT      /product= "Mature human DME-4"
FT      1..117
FT      /*tag= d
FT      mat_peptide
FT      118..1743
FT      /*tag= e
FT      /product= "Mature human DME-4"
FT      WO200246426-A2.
FT      XX
FT      XX
FT      PD      13-JUN-2002.
FT      XX
FT      PF      04-DEC-2001; 2001WO-US47429.
FT      XX
FT      PR      08-DEC-2000; 2000US-254308P.
FT      PR      15-DEC-2000; 2000US-256189P.
FT      PR      21-DEC-2000; 2000US-257113P.
FT      PR      19-JAN-2001; 2001US-262706P.
FT      PR      02-FEB-2001; 2001US-266020P.
FT      XX
FT      PA      (INCY-) INCYTE GENOMICS INC.
FT      XX
FT      PI      Sanjanwala MM, Yao MG, Au-young J, Baughn MR, Arvizu C, Ring HZ;
FT      PI      Lee EA, Ding L, Hafalia AJA, Tang YT, Yue H,
FT      PI      Lu DAM, Lal PG, Warren BA, Yang J, Wallia NK, Nguyen DB;
FT      PI      Gandhi AR, Lu Y, Ison CH;
FT      XX
FT      DR      WPI; 2002-519668/55.
FT      DR      P-FSDB; AAE25019.
FT      XX
FT      PT      Novel human drug metabolizing polypeptide, useful in diagnosis,
FT      PT      prevention or treatment of autoimmune/inflammatory, cell proliferative,
FT      PT      neurological, developmental, endocrine, metabolic and gastrointestinal
FT      PT      disorders .
FT      XX
FT      PS      Claim 72; Page 161-162; 169pp; English.
FT      XX
FT      CC      The invention relates to an isolated human drug metabolising enzyme (DME)
FT      CC      and its nucleotide. DME is useful for diagnosing, treating or preventing
FT      CC      disorders associated with aberrant expression of DME, where the disorders
FT      CC      are selected from autoimmune/inflammatory disorder such as acquired
FT      CC      immunodeficiency syndrome (AIDS), asthma, atherosclerosis, psoriasis,
FT      CC      uveitis; a cell proliferative disorder such as arteriosclerosis,
FT      CC      cirrhosis, hepatitis, and cancer; a neurological disorder such as
FT      CC      Alzheimer's disease, Huntington's disease, dementia, Parkinson's disease;
FT      CC      a developmental disorder such as renal tubular acidosis, epilepsy,
FT      CC      anaemia; an endocrine disorder such as adenoma, thrombosis and
FT      CC      infections; an eye disorder such as conjunctivitis, glaucoma, cataract;
FT      CC      metabolic disorder such as cystic fibrosis, diabetes and goitre; a
FT      CC      gastrointestinal disorder such as anorexia, peptic ulcer; and liver
FT      CC      disorders. DME is useful in a number of drug screening techniques and to
FT      CC      analyse the proteome of a tissue or cell type. The invention is useful
FT      CC      for creating knock-in humanised animals or transgenic animals to model
FT      CC      human diseases, in somatic or germline gene therapy, to generate a
FT      CC      transcript image of a tissue or cell type, for detecting differences in
FT      CC      the chromosomal location due to translocation, inversion, etc. among
FT      CC      normal, carrier or affected individuals, and as hybridisation probes for
FT      CC      mapping naturally occurring genomic sequences. The present sequence is
FT      CC      human DME-4 cDNA.
FT      XX
FT      SQ      Sequence 2745 BP; 704 A; 612 C; 737 G; 692 T; 0 other;
FT      XX
      Query Match
      Best Local Similarity 100.0%; Score 1746; DB 24; Length 2745;
      Matches 1746; Conservative 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
      1 ATGTGGGGGCGACGCGCGCGCGCGCTGCCCGGGAACTCGCGCGGAACTCGCGCGGCGCGGAGGCG 60
      1 ATGTGGGGGCGACGCGCGCGCGCGCTGCCCGGGAACTCGCGCGGAACTCGCGCGGCGCGGAGGCG 60
      61 CTGTTGGTGTCTCTGGCGGCTACTGGCGTGGCGGGCTGGGCTGGTCTCGGGCGGCGAG 120
      61 CTGTTGGTGTCTCTGGCGGCTACTGGCGTGGCGGGCTGGGCTGGTCTCGGGCGGCGAG 120

```


Db 441 TGAACCCGCTGTGCAAGAGAGAAATATGATATGATAATTTGCCAGGACATCTGTT 500
Qy 421 ATCATAGCATTTTATATAGAGCCGTGCTCAACTCTCTCGGACAGTTTACAGTGTCTTT 480
Db 501 ATCATAGCATTTTATATAGAGCCGTGCTCAACTCTCTCGGACAGTTTACAGTGTCTTT 560
Qy 481 GAGACATCCCCGGGATATCTCTAGAGAGAGTGTATCTTTGTAGATGACTACAGTGTATGA 540
Db 561 GAGACATCCCCGGGATATCTCTAGAGAGAGTGTATCTTTGTAGATGACTACAGTGTATGA 620
Qy 541 GAGACATGGAAGAGCGCTTGGCCAAATGAGCTTTTGGGACGTGCCAAAGGTGCGCTGATC 600
Db 621 GAGCACCTGGAAGAGCGCTTGGCCAAATGAGCTTTTGGGACGTGCCAAAGGTGCGCTGATC 680
Qy 601 CGCCGCAACAGAGAGAGCGCTTGGGAGAGCGGCTGCTGGGGGGCTCTGCGCGGAGG 660
Db 681 CGCCGCAACAGAGAGAGCGCTTGGGAGAGCGGCTGCTGGGGGGCTCTGCGCGGAGG 740
Qy 661 GCGGATGTTCTGACCTTCTCGGACTGTCACTGTGAGTGCACGAAGGTGCTGGAGCGG 720
Db 741 GCGGATGTTCTGACCTTCTCGGACTGTCACTGTGAGTGCACGAAGGTGCTGGAGCGG 800
Qy 721 CTGCTGAGAGAGATCCATGAAGAGAGTCCGCACTGTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db 801 CTGCTGAGAGAGATCCATGAAGAGAGTCCGCACTGTGCTGCTGCTGCTGCTGCTGCTGCTG 860
Qy 781 GACTGGACACCTTTCGATATCTGGGAACTCCCGGGAGCCCGAGATCGGCGTTTCGAC 840
Db 861 GACTGGACACCTTTCGATATCTGGGAACTCCCGGGAGCCCGAGATCGGCGTTTCGAC 920
Qy 841 TGGAGGCTGGTGTTCACGTGTCACAGTTCCTGAGAGGAGAGATACGATGCAATCC 900
Db 921 TGGAGGCTGGTGTTCACGTGTCACAGTTCCTGAGAGGAGAGATACGATGCAATCC 980
Qy 901 CCCGTCGATGTCATCAGTCTCCAAACATGCTGCTGGGCTGTTTGTGTGAGTAAGAAA 960
Db 981 CCCGTCGATGTCATCAGTCTCCAAACATGCTGCTGGGCTGTTTGTGTGAGTAAGAAA 1040
Qy 961 TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAACCTC 1020
Db 1041 TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAACCTC 1100
Qy 1021 GAAATTTCTTTAGATCTGCGAGTGTGGGGTCTTGGGAAACACACCCATGTTCCCAT 1080
Db 1101 GAAATTTCTTTAGATCTGCGAGTGTGGGGTCTTGGGAAACACACCCATGTTCCCAT 1160
Qy 1081 GTTGGCCATGTTTCCCAAGCAAGCTCTCTACTCCCGCAACAAAGGCTCTGGCCAAAGT 1140
Db 1161 GTTGGCCATGTTTCCCAAGCAAGCTCTCTACTCCCGCAACAAAGGCTCTGGCCAAAGT 1220
Qy 1141 GTTGGTCAGCTGAGTATGATGATGAATTTAAAGGCTCTACTACCATCGCAACCCC 1200
Db 1221 GTTGGTCAGCTGAGTATGATGATGAATTTAAAGGCTCTACTACCATCGCAACCCC 1280
Qy 1201 CGTGGCCCTTGGAACTTTTGGGATGTGACAGAGAGAGAGAGCTCCGGGCAAGCTC 1260
Db 1281 CGTGGCCCTTGGAACTTTTGGGATGTGACAGAGAGAGAGAGCTCCGGGCAAGCTC 1340
Qy 1261 CAGTGTAAGAATTCAAGTGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGAG 1320
Db 1341 CAGTGTAAGAATTCAAGTGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGAG 1400
Qy 1321 GACAGGCTGCTTCTTGGATGCTCCAGAACAAAGACTTAAAGACTACTGCTTTGAC 1380
Db 1401 GACAGGCTGCTTCTTGGATGCTCCAGAACAAAGACTTAAAGACTACTGCTTTGAC 1460
Qy 1381 TATAACCTCCCGATGAAACACAGATTGTGGGACACCAAGGCTATTCTGCTCTGTCAT 1440
Db 1461 TATAACCTCCCGATGAAACACAGATTGTGGGACACCAAGGCTATTCTGCTCTGTCAT 1520
Qy 1441 GGGATGGGCCAGAAATCAGTTTTTTCAGATACAGTCCCGAGAAAGAAATACGCTATAACACC 1500
Db 1521 GGGATGGGCCAGAAATCAGTTTTTTCAGATACAGTCCCGAGAAAGAAATACGCTATAACACC 1580

Qy 1501 CACCAGCCTGAGGCTGCATTGCTGGAGAGCAGGAATGATACCCCTTATCATGCATCTC 1560
Db 1581 CACCAGCCTGAGGCTGCATTGCTGGAGAGCAGGAATGATACCCCTTATCATGCATCTC 1640
Qy 1561 TCGGAAGAACTCCCCAGAGAAATCAGAGTTTCATCTTGCAGGAGGATGATCTTTATTT 1620
Db 1641 TCGGAAGAACTCCCCAGAGAAATCAGAGTTTCATCTTGCAGGAGGATGATCTTTATTT 1700
Qy 1621 CACGAACAGTCCCAAGAAATGTGTCCAGGCTGCGAGGAGAGAGTCCGAGTGACAGTTTCGTT 1680
Db 1701 CACGAACAGTCCCAAGAAATGTGTCCAGGCTGCGAGGAGAGAGTCCGAGTGACAGTTTCGTT 1760
Qy 1681 CCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1740
Db 1761 CCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1820
Qy 1741 TTATGA 1746
Db 1821 TTATGA 1826

RESULT 3
AAL26522/c
ID AAL26522 standard; cDNA; 2290 BP.
XX AAL26522;
AC AC
XX AC
DT 07-DEC-2001 (first entry)
XX 07-DEC-2001 (first entry)
DE Human breast cancer expressed polynucleotide 18979.
KW Human; breast cancer; cell marker; cytostatic; ss.
XX Homo sapiens.
XX WO200151628-A2.
XX 19-JUL-2001.
XX 10-JAN-2001; 2001WO-US00798.
XX 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Lillie J, Xu Y, Wang Y, Steinmann K;
PI WPI; 2001-451856/48.
XX New peptide useful as a marker for the diagnosis of breast cancer -
XX Claim 1; Page 3534-3535; 3695pp; English.
PS The invention relates to human breast cancer expressed polynucleotides
XX (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterizing treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
XX activity.

Query Match 99.1%; Score 1730.8; DB 22; Length 2290;

Best Local Similarity 99.8%; Pred. No. 0;			
Matches 1743; Conservative 0; Mismatches 2; Indels 1; Gaps 1;			
Qy	1	ATGTGGGGGCGACCGGGCGCGCGCTGCGCGGGAACCTGCGGGCGCGCGCGCGCGCGCG	60
Db	2212	ATGTGGTGGGCGACCGGGCGCGCGCTGCGCGGGAACCTGCGGGCGCGCGCGCGCGCGCG	2153
Qy	61	CTGTGGTGTCTTGGGGCTACTGCGGTGGCGGGGCTGGGCTGGGCTGGGCTGGGCTGGG	120
Db	2152	CTGTGGTGTCTTGGGGCTACTGCGGTGGCGGGGCTGGGCTGGGCTGGGCTGGGCTGGG	2093
Qy	121	CGTGGGGCGGGGCGGGGCTGCGGAGCGGGGACCCCGCGGACCCCGCGCGCGCGCGCGCG	180
Db	2092	CGTGGGGCGGGGCGGGGCTGCGGAGCGGGGACCCCGCGGACCCCGCGCGCGCGCGCGCG	2033
Qy	181	CGGAGCGGGTCAATGCG	240
Db	2032	CGGAGCGGGTCAATGCG	1973
Qy	241	CGGGTGGGGTGCAGCTGCGAGGGCGAGAGCTGCGGCTGCAGAGGAGAGCGTGGCGCTG	300
Db	1972	CGGGTGGGGTGCAGCTGCGAGGGCGAGAGCTGCGGCTGCAGAGGAGAGCGTGGCGCTG	1913
Qy	301	CACGAGATTAACTACTACCTCAGCGACCGCATCTCACTGCGCGCGCGCGCGCGCGCGCG	360
Db	1912	CACGAGATTAACTACTACCTCAGCGACCGCATCTCACTGCGCGCGCGCGCGCGCGCGCG	1853
Qy	361	TGGAACCGCGTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCCGAGGACATCTGTT	420
Db	1852	TGGAACCGCGTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCCGAGGACATCTGTT	1793
Qy	421	ATCATAGCATTTTAAATGAAGCTGTGTCACCTCTCTCGGACAGTTTACAGTGTCTTT	480
Db	1792	ATCATAGCATTTTAAATGAAGCTGTGTCACCTCTCTCGGACAGTTTACAGTGTCTTT	1733
Qy	481	GAGACATCCCGGATCTCTGCTAGAGAGATGATCTTGTAGATGATGATGATGATGATGATG	540
Db	1732	GAGACATCCCGGATCTCTGCTAGAGAGATGATCTTGTAGATGATGATGATGATGATGATG	1673
Qy	541	GAGCACCCTGAAGAGCGCTTTGGGCAATGAGCTTTGGGCACTGCCCAAGGTGGCGCTGATC	600
Db	1672	GAGCACCCTGAAGAGCGCTTTGGGCAATGAGCTTTGGGCACTGCCCAAGGTGGCGCTGATC	1613
Qy	601	CGGGCCAAAGAGAGAGCGCTGCTGCGAGCGCGCTGCTGGGGCGCTGCTGGGGCGAGG	660
Db	1612	CGGGCCAAAGAGAGAGCGCTGCTGCGAGCGCGCTGCTGGGGCGCTGCTGGGGCGAGG	1553
Qy	661	GGCGATGTTCTGACCTTCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAG	720
Db	1552	GGCGATGTTCTGACCTTCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAG	1493
Qy	721	CTGCTGCAGAGGATCCATGAAGAGGAGTGGCGAGTGGTGGCGGCTGATTTGATGATGATC	780
Db	1492	CTGCTGCAGAGGATCCATGAAGAGGAGTGGCGAGTGGTGGCGGCTGATTTGATGATGATC	1433
Qy	781	GACTGGAACACCTTCGGAATACCTGGGAACTCCGGGAGCCCGAGATCGGCGGTTTCGAC	840
Db	1432	GACTGGAACACCTTCGGAATACCTGGGAACTCCGGGAGCCCGAGATCGGCGGTTTCGAC	1373
Qy	841	TGGAGGCTGGTGTTCAGTGGGCAACAGTTCCTGAGGGGAGAGGATACGGATGCAATCC	900
Db	1372	TGGAGGCTGGTGTTCAGTGGGCAACAGTTCCTGAGGGGAGAGGATACGGATGCAATCC	1313
Qy	901	CCCGTGCATGTCATAGGTCTCAACAAATGGCTGGTGGGCTGTTGCTGTGAGTAAAGAA	960
Db	1312	CCCGTGCATGTCATAGGTCTCAACAAATGGCTGGTGGGCTGTTGCTGTGAGTAAAGAA	1253
Qy	961	TATTTGAAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAAACCTC	1020
Db	1252	TATTTGAAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAAACCTC	1193
Qy	1021	GAATTTTCTTTAGGATCTGGGAGTGTGGGGTCTTGGAACACACCCATCTTCCCAT	1080

RESULT 4

AAL26538/c

ID AAL26538 standard; cDNA; 2290 bp.

XX

AC AAL26538;

XX

DT 07-DEC-2001 (first entry)

XX

DE Human breast cancer expressed polynucleotide 18995.

XX

KW Human; breast cancer; cell marker; cytosolic; ss.

XX

OS Homo sapiens.

XX

PN WO200151628-A2.

XX

XX 19-JUL-2001.

XX

PF 10-JAN-2001; 2001WO-US00798.

XX

PR 14-JAN-2000; 2000US-0176077.

PR 14-MAR-2000; 2000US-0189167.

Db	1192	GAATTTTCTTTAGGATCTGGCAGTGTGGGGTCTTGAAAAACACACCCATGTTCCCAT	1133
Qy	1081	GTTGGGCATGTTTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCACAGT	1140
Db	1132	GTTGGGCATGTTTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCACAGT	1073
Qy	1141	GTTCTGTGAGCTGAAATGATGGATGGAATTTAAAGAGCTCTACTACCATCGCAACCCC	1200
Db	1072	GTT-GTGACAGCTGAAATGATGGATGGAATTTAAAGAGCTCTACTACCATCGCAACCCC	1014
Qy	1201	CGTGGCCGCTTGAACCTTTTGGGGATGTGACAGAGAGGAGCAGCTCCGGGCAAGCTC	1260
Db	1013	CGTGGCCGCTTGAACCTTTTGGGGATGTGACAGAGAGGAGCAGCTCCGGGCAAGCTC	954
Qy	1261	CAGTGTAAAGACTTCAAGTGGTCTTGGGAGCTGTGTATCCAGAACTGCAATGTGCCTGAG	1320
Db	953	CAGTGTAAAGACTTCAAGTGGTCTTGGGAGCTGTGTATCCAGAACTGCAATGTGCCTGAG	894
Qy	1321	GACAGGCTGGCTTCTTGGGATGCTCCAGAAACAAAGGACTAAACAGACTACTGCTTTGAC	1380
Db	893	GACAGGCTGGCTTCTTGGGATGCTCCAGAAACAAAGGACTAAACAGACTACTGCTTTGAC	834
Qy	1381	TATAACCTCCCGATGAAACCCAGATTTGGGGACACCAAGTCTTGTACCTCTGTCTAT	1440
Db	833	TATAACCTCCCGATGAAACCCAGATTTGGGGACACCAAGTCTTGTACCTCTGTCTAT	774
Qy	1441	GGATGGGCGCAGAAATCAGTCTTTTCCGAGTACACGTCCAGAAAGAAATACGCTATAACAC	1500
Db	773	GGATGGGCGCAGAAATCAGTCTTTTCCGAGTACACGTCCAGAAAGAAATACGCTATAACAC	714
Qy	1501	CACGAGCTGAGGGCTGCAATGCTGTGGAAGCAGGAATGGAATACCTTTATCATGCACTC	1560
Db	713	CACGAGCTGAGGGCTGCAATGCTGTGGAAGCAGGAATGGAATACCTTTATCATGCACTC	654
Qy	1561	TGCGAAGAACTGCCCCAGAGAAATCAGAGTTCATCTTGAGAGAGGATGGAATCTTTATTT	1620
Db	653	TGCGAAGAACTGCCCCAGAGAAATCAGAGTTCATCTTGAGAGAGGATGGAATCTTTATTT	594
Qy	1621	CACGAAACAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAGAGTCCGAGTTCAGTTTCGTT	1680
Db	593	CACGAAACAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAGAGTCCGAGTTCAGTTTCGTT	534
Qy	1681	CCACTTTACGAGACTGCCAACCTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG	1740
Db	533	CCACTTTACGAGACTGCCAACCTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG	474
Qy	1741	TTATGA 1746	
Db	473	TTATGA 468	

Db 533 CCACTCTTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATG 474

Qy 1741 TTATGA 1746
|||||

Db 473 TTATGA 468

RESULT 5

AAC98209

ID AAC98209 standard; cDNA; 1259 BP.

XX AAC98209;

AC AC

DT 09-MAR-2001 (first entry)

XX Human colon cancer antigen nucleotide sequence SEQ ID NO:219.

DE Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;

KW immunomodulatory; muscular; gynaecological; gastrointestinal;

KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;

KW neural disorder; immune system disorder; muscular disorder;

KW reproductive disorder; gastrointestinal disorder; renal disorder;

KW infectious disease; cardiovascular disorder; ss.

XX Homo sapiens.

OS

XX WO20005351-A1.

PN

XX 21-SEP-2000.

PD

XX 08-MAR-2000; 2000WO-US05883.

PF

XX 12-MAR-1999; 99US-0124270.

PR

XX (HUMA-) HUMAN GENOME SCI INC.

PA

XX Rosen CA, Ruben SM;

PI

XX WPI; 2000-587534/55.

DR

XX P-PSDB; AAB53452.

XX

PT Colon cancer associated gene sequences, referred to as colon cancer

PT antigens, useful for the treatment, prevention, and diagnosis of colon

PT disorders such as colon cancer -

XX

PS Claim 1; Page 642; 2104pp; English.

XX

CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,

CC called human colon cancer antigens, given in AAB53234 to AAB54006. The

CC human colon cancer antigens can have cytostatic, cardioactive, muscular;

CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,

CC vulnerary, nephrotropic, antiinfective and antibacterial activities, and

CC can be used in gene therapy. The colon cancer antigen polynucleotides,

CC proteins and antibodies to the proteins are useful for the prevention,

CC treatment and diagnosis of colon disorders, such as colon cancer. The

CC polynucleotides may be used in diagnostics and research, such as for

CC chromosome identification, and as hybridisation probes. The proteins

CC may also be used to prevent diseases such as neural disorders, immune

CC system disorders, muscular disorders, reproductive disorders,

CC gastrointestinal disorders, wounds, renal disorders, infectious

CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and

CC AAB54007 represent sequences used in the exemplification of the present

CC invention.

XX

SQ Sequence 1259 BP; 284 A; 310 C; 383 G; 278 T; 4 other;

Query Match 69.2%; Score 1208.4; DB 21; Length 1259;

Best Local Similarity 98.4%; Pred. No. 1.4e-238;

Matches 1226; Conservative 4; Mismatches 15; Indels 1; Gaps 1;

Qy 212 CGGGAACGGCTGGGCGCGGCGGCGGCTGCGGCTGCGAGCTCGAGGCGGCGAGGAGC 271
|||||

Db 7 CGGTCCGGAACCGCTGGGCGGAGCGAGCGGTGCGGCTGCGAGCTGCGAGGCGAGAGC 66

Qy 272 TGCGGCTGCAGGAGAGCGTGCGGCTGCAACAGATTAACTACTCTACCTCAGCGACCGCA 331
|||||

Db 67 TGCGGCTGCAGGAGAGCGTGCGGCTGCAACAGATTAACTACTCTACCTCAGCGACCGCA 126
|||||

Qy 332 TCTCACTGACCGCGGCTGCGGAGCGGTGGAAACCGGCTGTGCAAGAGAAGAAATATG 391
|||||

Db 127 TCTCACTGACCGCGGCTGCGGAGCGGTGGAAACCGGCTGTGCAAGAGAAGAAATATG 186
|||||

Qy 392 ATTATGATTAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCCCTGGTCAA 451
|||||

Db 187 ATTATGATTAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCCCTGGTCAA 246
|||||

Qy 452 CTCTCTTCGACAGTTTACAGTGTCTCTGAGACATCCCCGGATATCTCTGTAGAAGAAG 511
|||||

Db 247 CTCTCTTCGACAGTTTACAGTGTCTCTGAGACATCCCCGGATATCTCTGTAGAAGAAG 306
|||||

Qy 512 TGATCCTTTAGATGATACAGTGTAGAGAGACCTGAAAGGAGCGTTGGCCCAATGAGC 571
|||||

Db 307 TGATCCTTTAGATGATACAGTGTAGAGAGACCTGAAAGGAGCGTTGGCCCAATGAGC 366
|||||

Qy 572 TTTTCGGGACTGCCCAAGGTGCGCTGATCCGCGCCCAACAGAGAGAGGCGCTGCTGCGAG 631
|||||

Db 367 TTTTCGGGACTGCCCAAGGTGCGCTGATCCGCGCCCAACAGAGAGAGGCGCTGCTGCGAG 426
|||||

Qy 632 CCCGGCTGCTGGGCGGCTCTGCGCGAGGGGCGATGTTCTGACCTTCTGAGACTGTCACT 691
|||||

Db 427 CCCGGCTGCTGGGCGGCTCTGCGCGAGGGGCGATGTTCTGACCTTCTGAGACTGTCACT 486
|||||

Qy 692 GTGAGTGCACGAAGGTGCTGGAGCGGCTGTGACAGAGATCCATGAAGAGAGTCCG 751
|||||

Db 487 GTGAGTGCACGAAGGTGCTGGAGCGGCTGTGACAGAGATCCATGAAGAGAGTCCG 545
|||||

Qy 752 CAGTGTGTGCCCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 811
|||||

Db 546 CAGTGTGTGCCCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 605
|||||

Qy 812 CCGGGAGAGCCCGAGATCGGCGGTTTCGACTGGAGGCTGGTGTTCACGTGGCACACAGTTC 871
|||||

Db 606 CCGGGAGAGCCCGAGATCGGCGGTTTCGACTGGAGGCTGGTGTTCACGTGGCACACAGTTC 665
|||||

Qy 872 CTGAGAGGAGAGGATACGAGTCAATCCCCCGTGCATGTCATGAGTCTTCAACAAATGG 931
|||||

Db 666 CTGAGAGGAGAGGATACGAGTCAATCCCCCGTGCATGTCATGAGTCTTCAACAAATGG 725
|||||

Qy 932 CTGGTGGGCTGTTGCTGTGAGTAAGAAATATTTGAATATCTGGGGTCTTATGATACAG 991
|||||

Db 726 CTGGTGGGCTGTTGCTGTGAGTAAGAAATATTTGAATATCTGGGGTCTTATGATACAG 785
|||||

Qy 992 GAATGGAAGTTTGGGAGGAGAGAAACCTCGAAATTTTCTTTAGGATCTGGCAGTGTGGTG 1051
|||||

Db 786 GAATGGAAGTTTGGGAGGAGAGAAACCTCGAAATTTTCTTTAGGATCTGGCAGTGTGGTG 845
|||||

Qy 1052 GGGTTCTGGAACACACCCATGTTCCATGTTGGCATGTTTCCCAAGCAAGTCCCT 1111
|||||

Db 846 GGGTTCTGGAACACACCCATGTTCCATGTTGGCATGTTTCCCAAGCAAGTCCCT 905
|||||

Qy 1112 ACTCCCGCACRAGGCTCTGGCCAAACAGTGTTCGTGCGAGCTGAATGATGATGATGATGAT 1171
|||||

Db 906 ACTCCCGCACRAGGCTCTGGCCAAACAGTGTTCGTGCGAGCTGAATGATGATGATGATGAT 965
|||||

Qy 1172 TTAAGAGGCTCTACTACCATCGCAACCCCGCTGCGGCTTGGAAACCTTTTGGGAGTGTGA 1231
|||||

Db 966 TTAAGAGGCTCTACTACCATCGCAACCCCGCTGCGGCTTGGAAACCTTTTGGGAGTGTGA 1025
|||||

Qy 1232 CAGAGAGGAGAGGCTCCGCGCAACAGCTCCAGTGTAAAGACTTCAAGTGGTCTTGGAGA 1291
|||||

Db 1026 CAGAGAGGAGAGGCTCCGCGCAACAGCTCCAGTGTAAAGACTTCAAGTGGTCTTGGAGA 1085
|||||

Qy 1292 CTGTGATCCAGAACTGCATGTCCTGAGGACAGGCGCTGCTTCTTCCGGATGCTCAGA 1351
|||||

Db 1086 CTGTGATCCAGAACTGCATGTCCTGAGGACAGGCGCTGCTTCTTCCGGATGCTCAGA 1145
|||||

Db 1021 CTTGG 1025

RESULT 7
ABV23465/c
ID ABV23465 standard; cDNA; 1352 BP.
XX AC
XX ABV23465;
XX
XX
XX 16-SEP-2002 (first entry)
XX
XX Human prostate expression marker cDNA 23456.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
XX OS
XX WO200160860-A2.
XX
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05171.
XX
XX 17-FEB-2000; 2000US-183319P.
XX PR
XX 16-MAR-2000; 2000US-189862P.
XX PR
XX 23-MAY-2000; 2000US-207454P.
XX PR
XX 09-JUN-2000; 2000US-211314P.
XX PR
XX 18-JUL-2000; 2000US-219007P.
XX PR
XX 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PA
XX Schlegel R, Endege WO, Monahan JE;
XX PI
XX WPI; 2001-662795/76.
XX DR
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX Claim 1; Page 4286-4287; 11750pp; English.
XX PS
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 1352 BP; 340 A; 335 C; 299 G; 368 T; 10 other;
XX SQ

Query Match 47.0%; Score 821.2; DB 23; Length 1352;
Best Local Similarity 99.5%; Pred. No. 4e-159;
Matches 834; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
XX
XX 909 TGTCTCAGCTCCACATGCTGGTGGGCTGTTTGGCTGTGAGTAAGAAATTTTGA 968
XX
XX 1304 TGCATGTGCTCCACATGCTGGTGGGCTGTTTGGCTGTGAGTAAGAAATTTTGA 1245
XX
XX 969 ATATCTGGGCTTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTC 1028
XX
XX 1244 ATATCTGGGCTTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTC 1185
XX

QY 1029 CTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAAACACACCCCATGTTCCCATGTTGGCCA 1088
Db 1184 CTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAAACACACCCCATGTTCCCATGTTGGCCA 1125
QY 1089 TGTTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAAGTGTTCGTGC 1148
Db 1124 TGTTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAAGTGTTCGTGC 1066
QY 1149 AGCTGAAGTATGGATGGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGCCTCG 1208
Db 1065 AGCTGAAGTATGGATGGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGCCTCG 1006
QY 1209 CTTTGGAACTTTTGGGATGTGACAGAGAGAGAGCTCCGGGCAAGCTCAGTGTAA 1268
Db 1005 CTTTGGAACTTTTGGGATGTGACAGAGAGAGAGCTCCGGGCAAGCTCAGTGTAA 946
QY 1269 AGACTTCAAGTGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCC 1328
Db 945 AGACTTCAAGTGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCC 886
QY 1329 TGGCTTCTTCGGGATGCTCCAGAACAAAGGACTTAACAGACTACTGCTTTGACTATAACCC 1388
Db 885 TGGCTTCTTCGGGATGCTCCAGAACAAAGGACTTAACAGACTACTGCTTTGACTATAACCC 826
QY 1389 TCCCGATGAAAACAGATTTGTGGACACAGGTCATTCTGTACCTCTGTCTGATGGATGGG 1448
Db 825 TCCCGATGAAAACAGATTTGTGGACACAGGTCATTCTGTACCTCTGTCTGATGGATGGG 766
QY 1449 CCAGAACTAGTTTTCAGTACACGCTCCAGAAAGAAATACGCTATTAACACCCACAGGCC 1508
Db 765 CCAGAACTAGTTTTCAGTACACGCTCCAGAAAGAAATACGCTATTAACACCCACAGGCC 706
QY 1509 TGAGGCTGCAATGCTGTGGAAGCAGGAATGGAATACCTTTATCATGCTCTCTGCGAAGA 1568
Db 705 TGAGGCTGCAATGCTGTGGAAGCAGGAATGGAATACCTTTATCATGCTCTCTGCGAAGA 646
QY 1569 AACTGCCCCAGAGAATCAGAAATTCATCTTTGAGGAGATGGATCTTTATTCACGAACA 1628
Db 645 AACTGCCCCAGAGAATCAGAAATTCATCTTTGAGGAGATGGATCTTTATTCACGAACA 586
QY 1629 GTCACAGAAATGTCCAGGCTCGAGGAGGAGTGCAGTGCAGATTCGTTCCACTCTT 1688
Db 585 GTCACAGAAATGTCCAGGCTCGAGGAGGAGTGCAGTGCAGATTCGTTCCACTCTT 526
QY 1689 ACAGAGCTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 1746
Db 525 ACAGAGCTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 468

RESULT 8
ABV29323/c
ID ABV29323 standard; cDNA; 1352 BP.
XX
XX AC ABV29323;
XX
XX DT 16-SEP-2002 (first entry)
XX
XX DE Human prostate expression marker cDNA 29314.
XX
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX PN WO200160860-A2.
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05171.
XX
XX 17-FEB-2000; 2000US-183319P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -
XX Claim 1; Page 6275; 11750pp; English.
XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX Sequence 1352 BP; 340 A; 335 C; 299 G; 368 T; 10 other;
Query Match 47.0%; Score 821.2; DB 23; Length 1352;
Best Local Similarity 99.5%; Pred. No. 4e-159;
Matches 834; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 909 TGCTATGAGTCTCCAAAGTGGTGGGCTGTTCTGTGAGTAAGAATATTTGA 968
DB 1304 TGCCATGTTGGTCTCCAAAGTGGTGGGCTGTTCTGTGAGTAAGAATATTTGA 1245
QY 969 ATATCTGGGCTTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTC 1028
DB 1244 ATATCTGGGCTTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTC 1185
QY 1029 CTTTAGGATCTGGCAGTGTGGGTTCTGGAACACACACCATGTTCCCATGTTGGCCA 1088
DB 1184 CTTTAGGATCTGGCAGTGTGGGTTCTGGAACACACACCATGTTCCCATGTTGGCCA 1125
QY 1089 TGTTTTCCTCCAGCAAGCTCCCTACTCCGCAACAGGCTCTGGCAACAGTGTTCGTGC 1148
DB 1124 TGTTTTCCTCCAGCAAGCTCCCTACTCCGCAACAGGCTCTGGCAACAGTGTTCGTGC 1066
QY 1149 AGCTGAAGTATGGATGGAATTTAAAGAGCTCTACTACCATCGCAACCCCTGGCCCG 1208
DB 1065 AGCTGAAGTATGGATGGAATTTAAAGAGCTCTACTACCATCGCAACCCCTGGCCCG 1006
QY 1209 CTTGGAACCTTTTGGGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAAGCTCCAGTGTAA 1268
DB 1005 CTTGGAACCTTTTGGGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAAGCTCCAGTGTAA 946
QY 1269 AGACTTCAAGTGTCTTGGAGACTGTGTATCCAGAACTGATGCTGAGGACAGGCC 1328
DB 945 AGACTTCAAGTGTCTTGGAGACTGTGTATCCAGAACTGATGCTGAGGACAGGCC 886
QY 1329 TGGCTTCTTGGGATGCTCCGAACAAAGGACTACAGACTACTCTTGTACTATAACC 1388
DB 885 TGGCTTCTTGGGATGCTCCGAACAAAGGACTACAGACTACTCTTGTACTATAACC 826
QY 1389 TCCCGATGAACACAGATTTGGGACACAGGTCAATCTGTACTCTCTCATGGATGGG 1448
DB 825 TCCCGATGAACACAGATTTGGGACACAGGTCAATCTGTACTCTCTCATGGATGGG 766

1449 CCAGAAATCAGTTTTTTCGAGTACACGTCACAGAAAGAAATACGCTATAACACCCAGGCC 1508
765 CCAGAAATCAGTTTTTTCGAGTACACGTCACAGAAAGAAATACGCTATAACACCCAGGCC 706
1509 TGAGGGCTGCATTGCTGTGGAAGCAAGAAATGGAATACCTTTATCATGATCTCTGCGAGA 1568
705 TGAGGGCTGCATTGCTGTGGAAGCAAGAAATGGAATACCTTTATCATGATCTCTGCGAGA 646
1569 AACTGCCCCCAGAGAAATCAGAAATGTTTCATCTTGCAGGAGGATGGATCTTTATTTCCAGAAC 1628
645 AACTGCCCCCAGAGAAATCAGAAATGTTTCATCTTGCAGGAGGATGGATCTTTATTTCCAGAAC 586
1629 GTCCAGAAATGTTTCAGGCTCGAGGAGGAGGAGTGCAGTGCAGATTTTCCTTCCACTCTT 1688
585 GTCCAGAAATGTTTCAGGCTCGAGGAGGAGGAGTGCAGTGCAGATTTTCCTTCCACTCTT 526
1689 ACAGAGTGCACCAACTCGGATCATCAGAAATGGTCTTCAAAGAGCGCATGTTATGA 1746
525 ACAGAGTGCACCAACTCGGATCATCAGAAATGGTCTTCAAAGAGCGCATGTTATGA 468
RESULT 9
AAH33532
ID AAH33532 standard; cDNA; 1517 BP.
XX AC AAH33532;
XX 03-SEP-2001 (first entry)
DE Human colon cancer antigen encoding cDNA SEQ ID NO:588.
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW Colorectal carcinoma; ss.
XX Homo sapiens.
OS WO200122920-A2.
PN 05-APR-2001.
XX 28-SEP-2000; 2000WO-US26524.
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI; 2001-235357/24.
XX P-PSDB; AAG74101.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers -
Claim 1; Page 2640; 9803pp; English.
AAH32943 to AAH37195 and AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patient's own production of P. Additionally, N may be used to produce the colon cancer-associated Ps by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent sequences used in the exemplification of the

Db 779 CTTCTGGATGCTCCAGAACAAAGGACTAAACAGACTACTGCTTTGACTATAACCTCCCGA 838
Qy 1395 TGAATAACAGATTGCGGACACAGGTCATTCTCTGCTCTGCTATGCGGATGGGCCAGAA 1454
Db 839 TGAATAACAGATTGCGGACACAGGTCATTCTCTGCTCTGCTATGCGGATGGGCCAGAA 898
Qy 1455 TCAGTTTTCAGTACAGCTCCAGAAAGAAATACGCTATACACCCACAGCCTGAGGG 1514
Db 899 TCAGTTTTCAGTACAGCTCCAGAAAGAAATACGCTATACACCCACAGCCTGAGGG 958
Qy 1515 CTGCATTGCTGGAAGCAGGAATGGATACCTTTATCATGCTCTGCGAAGAACTGC 1574
Db 959 CTGCATTGCTGGAAGCAGGAATGGATACCTTTATCATGCTCTGCGAAGAACTGC 1018
Qy 1575 CCCAGAGATCAGAAATTCATCTTCAGGAGGATGGATCTTTATTCACGACAGTCCAA 1634
Db 1019 CCCAGAGATCAGAAATTCATCTTCAGGAGGATGGATCTTTATTCACGACAGTCCAA 1078
Qy 1635 GAAATGTCTCAGGCTCGAGAGAGGAGTGCAGTGACAGTTTCGTTCCACTCTTTACGAGA 1694
Db 1079 GAAATGTCTCAGGCTCGAGAGAGGAGTGCAGTGACAGTTTCGTTCCACTCTTTACGAGA 1138
Qy 1695 CTGCACCAACTCGGATCATCAGAAATGGTCTTCCAAAGAGCGCATGTTATGA 1746
Db 1139 CTGCACCAACTCGGATCATCAGAAATGGTCTTCCAAAGAGCGCATGTTATGA 1190

RESULT 11

AAA78414
ID AAA78414 standard; cDNA; 1517 BP.
XX AC AAA78414;
XX DT 20-NOV-2000 (first entry)
XX DE Human secreted protein gene 34 SEQ ID NO:44.
XX KW Human; secreted protein; cytostatic; antianaemic; antidiabetic;
XX KW antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
XX KW antiporotic; angiogenic; cardiant; anti-HIV; neutropic;
XX KW neutropic; antimicrobial; antiparkinsonian; cancer;
XX KW immune system disorder; angiogenesis; hyperproliferative disorder;
XX KW cardiovascular disorder; apoptosis; neurological disease;
XX KW infectious disease; wound healing; ss.
XX OS Homo sapiens.
XX PN WO20035937-A1.
XX PD 22-JUN-2000.
XX PF 16-DEC-1999; 99WO-US29950.
XX PR 17-DEC-1998; 98US-0112809.
XX PR 18-DEC-1998; 98US-0113006.
XX XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Ebner R, Rosen CA, Endress GA, Soppet DR, Ni J;
XX PI Duan DR, Moore PA, Shi Y, Lafleur DW, Olsen HS, Florence K;
XX XX
XX DR WPI; 2000-431566/37.
XX DR P-PSDB; AAB24470.
XX XX
XX PT Forty seven human nucleic acids encoding secreted proteins, useful in
XX PT the treatment, prevention and diagnosis of cancers, disorders of the
XX PT immune system, angiogenesis disorders, neurological diseases and
XX PT hyperproliferative disorders -
XX PS Claim 1; Page 465-466; 562pp; English.
XX CC The polynucleotide sequence given in AAA78381 to AAA78432 encode the

CC human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: cytostatic; antianaemic;
CC antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC antiarthritic; antiporotic; angiogenic; cardiant; anti-HIV;
CC neutropic; neutropic; antimicrobial; antiparkinsonian;
CC Human secreted protein polynucleotides, polypeptides, antagonists and/or
CC agonists may be useful in treating, preventing, and/or diagnosing other
CC diseases, disorders, and/or conditions such as: (a) cancers; (b)
CC disorders of the immune system; (c) angiogenesis disorders; (d)
CC hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
CC associated with increase apoptosis; (g) neurological diseases; and
CC (h) infectious diseases. They are also used to promote wound healing.
CC AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC exemplification of the present invention.

XX Sequence 1517 BP; 366 A; 360 C; 363 G; 424 T; 4 other;

Query Match 40.8%; Score 711.6; DB 21; Length 1517;
Best Local Similarity 99.9%; Pred. No. 1.3e-136;
Matches 711; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1035 GATCTGGCAGTGTGGTGGGTTCTTGGAAACACACCCATGTTCCCATGTTGGCCATGTTT 1094
Db 479 GATCTGGCAGTGTGGTGGGTTCTTGGAAACACACCCATGTTCCCATGTTGGCCATGTTT 538
Qy 1095 CCCAAGCAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGTCTTCTGTCAGCTGA 1154
Db 539 CCCAAGCAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGTCTTCTGTCAGCTGA 598
Qy 1155 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGA 1214
Db 599 AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGA 658
Qy 1215 ACCTTTGGGATGTGACAGAGGAGAGCTCCGGGACAGCTCCAGTGTAAAGACTT 1274
Db 659 ACCTTTGGGATGTGACAGAGGAGAGCTCCGGGACAGCTCCAGTGTAAAGACTT 718
Qy 1275 CAAGTGGTCTTCGAGACTGTGTATCCAGAACTGCTATGCTCCCTGAGGACAGGCTGCTT 1334
Db 719 CAAGTGGTCTTCGAGACTGTGTATCCAGAACTGCTATGCTCCCTGAGGACAGGCTGCTT 778
Qy 1335 CTTTCGGGATGCTCCAGAAACAAAGGACTTAACAGACTACTGCTTTGACTATAACCTCCGA 1394
Db 779 CTTTCGGGATGCTCCAGAAACAAAGGACTTAACAGACTACTGCTTTGACTATAACCTCCGA 838
Qy 1395 TGAACACAGATTGTGGGACACAGGTCATTCTGTACTCTCTCATCGGATGGCCAGAA 1454
Db 839 TGAACACAGATTGTGGGACACAGGTCATTCTGTACTCTCTCATCGGATGGCCAGAA 898
Qy 1455 TCAGTTTTCAGTACAGCTCCAGAAAGAAATACGCTATAACACCCACAGCCTGAGGG 1514
Db 899 TCAGTTTTCAGTACAGCTCCAGAAAGAAATACGCTATAACACCCACAGCCTGAGGG 958
Qy 1515 CTGCATTGCTGGAAGCAGGAATGGATACCTTTATCATGCTCTGCGAAGAACTGC 1574
Db 959 CTGCATTGCTGGAAGCAGGAATGGATACCTTTATCATGCTCTGCGAAGAACTGC 1018
Qy 1575 CCCAGAGATCAGAAATTCATCTTCAGGAGGATGGATCTTTATTCACGACAGTCCAA 1634
Db 1019 CCCAGAGATCAGAAATTCATCTTCAGGAGGATGGATCTTTATTCACGACAGTCCAA 1078
Qy 1635 GAAATGTCTCAGGCTCGAGAGGAGGATCGAGTGACAGTTTCGTTCCACTCTTTACGAGA 1694
Db 1079 GAAATGTCTCAGGCTCGAGAGGAGGATCGAGTGACAGTTTCGTTCCACTCTTTACGAGA 1138
Qy 1695 CTGCACCAACTCGGATCATCAGAAATGGTCTTCCAAAGAGCGCATGTTATGA 1746
Db 1139 CTGCACCAACTCGGATCATCAGAAATGGTCTTCCAAAGAGCGCATGTTATGA 1190

RESULT 12
ABX14962

```
ID ABX14962 standard; DNA; 1737 BP.
XX ABX14962;
XX
XX 13-MAR-2003 (first entry)
XX Human N-acetylgalactosaminyl transferase T4, GalNac T4, gene.
XX
XX Human; ds; gene; GalNac T4; N-acetylgalactosaminyl transferase T4;
KW Glycosylation; mucin 1; MUC1; vaccine; antiinflammatory; GalNac-T1;
KW GalNac-T2; GalNac-T3.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1737
XX /*tag= a
XX /product= "GalNac T4"
XX
XX US645220-B1.
XX
XX 15-OCT-2002.
XX
XX 21-DEC-1998; 98US-0217306.
XX
XX 21-DEC-1998; 98US-0217306.
XX (GLYC-) GLYCOZYM APS.
XX
XX Hassan FH, Clausen H, Bennett EP, Eisenkraetzer D, Gaetgens J;
XX WPI; 2003-147066/14.
XX P-PSDB; ABU07527.
XX
XX Glycosylating MUC1 acceptor substrate, by glycosylating substrate with
XX N-acetylgalactosaminyltransferase T1, GalNac-T2 or GalNac-T3, then with
XX human GalNac-T4 to glycosylate specific Ser, Thr residues in substrate
XX
XX Example 1; Column 9-12; 10pp; English.
XX
XX The invention relates to glycosylating a MUC1 (mucin 1) acceptor
XX substrate, comprising glycosylating the substrate with enzymatically
XX active N-acetylgalactosaminyltransferase (GalNac)-T1, GalNac-T2 or
XX GalNac-T3, or with GalNac capable of glycosylating MUC1 glycosylation
XX sites that can be glycosylated by GalNac-T1, GalNac-T2 or GalNac-T3, and
XX glycosylating the substrate with enzymatically active human GalNac-T4 to
XX glycosylate specific Ser, Thr positions in the MUC1 substrate.
XX The method is used for glycosylating an MUC1 acceptor substrate.
XX Glycosylated substrates are useful in preparation of vaccines and
XX antiinflammatory agents. GalNac-T4 exhibits a different substrate
XX specificity than previously characterised GalNac transferases. The
XX activity of GalNac-T4 is unique and specific to glycosylate specific
XX serine and threonine residues in MUC1 tandem repeat.
XX The present sequence is the human GalNac T4 gene.
XX
XX Sequence 1737 BP; 482 A; 354 C; 437 G; 464 T; 0 other;
XX
XX Query Match 28.4%; Score 496.2; DB 25; Length 1737;
XX Best Local Similarity 62.3%; Pred. No. 2e-92;
XX Matches 796; Conservative 0; Mismatches 478; Indels 3; Gaps 1;
XX
XX 220 GCGTGGCGCGGGCGAGCGGTGCGGTGCGAGCTGCAGGCGGAGGAGCTGCGGCTG 279
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 217 GCATTGGGGAGTGGGGGAAGCCAGCAAACTCCAGCTCAACGAGGATGAAGTGAAGCAG 276
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 280 CAGGAGAGAGCGTGGCGGTGACCAAGATTAACTTACTTACCGGACCGGATCTCACTG 339
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 277 CAAGAGAACTCAATTGAGAGATAGCCCATCAATATTACCTCACTGACAGGATTCCTCG 336
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 340 CACCGCGCTGCCGAGCGCTGGAAACCGCTGTGCAAGAGAGAAATATGATTATGAT 399
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 337 CATCGACACATAGAGGATAAAGAAATGATATGAGTGTAAAGTCCCAAGAGATTCAACTATAGG 396
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 400 AATTTGCCAGGACATCTGTTATCATAGCAATTTTATAATGAACCTGGTCAACTCTCCTT 459
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 397 ACACCTTCTACACACCTCTGTTATCATGCTTTTATAACGAAGCGTGGTGCATTTGCTC 456
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 460 CGGACAGTTTACAGTGTCTTGGAGATCCCGGATATCTCTAGNAGAAGTATCTT 519
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 457 CGTACCAATTCACAGTGTTTTAGAAACTTCTCTGCACTTCTTTGAAAGAGATCATCTG 516
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 520 GTAGATGACTACAGTGTATAGAGAGCACCTGAAGGAGCGCTTGGCCATAGAGCTTCCGGA 579
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 517 GTGATGACTTGTAGTGACAGAGTTTATTTGAAAGACAACTTGAACCTTACATCAGCAAT 576
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 580 CTGCCAAAGGTGCGCTGATCCCGCCAAACAAGAGAGAGCGCTGTGCGAGCCGCTG 639
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 577 CTTGATAGAGTACGCTTGATTAGGACCAATAAGCAGAGGGGCTGTTAGGCGCGCTG 636
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 640 CTGGGGCGCTCTGGGGGAGGGGGGATGTTCTGACCTTCTGAGATGTCTATGTGATGTC 699
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 637 AATTGGGGCCACTTTTCGCCACTGGGGAGCTCCTCACTTCTCTATTTGTCTACTGTGAGTGT 696
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 700 CACGAAGGGTGGCTGGAGCGCTGCTGCGAGAGATCCATGAAGAGGAGTCCGCGAGTGGT 759
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 697 AATTCGGTTGGCTGGACCGCTTTTGGAAAGGATTTGGGAGATATGAAACAGCAGTTGTG 756
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 760 TGCCCGGTGATTGATGCTGACTGGAAACACTTCCGAATACCTGGGGAATCTCCGGGGAG 819
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 757 TGTCTCTTATAGACACAAATTGGAATTAATTTTGAATTTCTATATATGACAGATAGGGAG 816
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 820 CCCAGATCGGCGGTTTCGACTGGAGGCTGCTTCACTGCGGACACAGTCTCTGAGAGG 879
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 817 CCCATGATTGGTGGGTTGACTGGCGCTTTAAACATTTTCAGTGGCATCTGTCCCCAAACAG 876
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 880 GAGAGATACGAGTGAATCCCGCTCGATGTCTCATCAGGTCTCCAACTAGGTGGTGGG 939
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 877 GAAAGGACAGCGGATATCAAGAAATTGACCCCATCAGATCACTACCATGGCTGGAGGA 936
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 940 CTGTTTCTGTGATGAAGAAATATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAA 999
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 937 CTGTTTCTGTGACGAAGAAATATTTTCACTACCTTGGAACTGATGACACAGGAATGGAA 996
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1000 GTTTGGGAGGAGAAAACCTCGAATTTTCTTTAGATCTGGGAGTGTGGTGGGTTCTG 1059
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 997 GTGTGGGAGGTGAAACCTTGAGCTGTCTTTTAGGGTGTGGCAGTGTGGTGGCAATTTG 1056
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1060 GAAACACACCCATGTTCCCATGTTGCGCATGTTTCCCAAGCAAGCTCCCTACTCCCGC 1119
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 1057 GAGATCCACCCGTGTTCCACGTGGGCGCATGTGTTCCCAAGCGGCGCACCATATGCTCGC 1116
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1120 AACAAAGGCTCTGGCCAAACAGTGTTCGTGACAGCTGAAGTATGGATGAATTTAAAGAG 1179
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 1117 CCCAATTTCTACAGAATACTGCTCGGGCAGCAGAAAGTTTGGATGGATGAATACAAAGAG 1176
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1180 CTCTACTACCATCGAAACCCCGCTGCCCGTGGAACTTTTGGGAGTGTGACAGAGAGG 1239
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 1177 CACTTCTCAATAGAAACCCCTCCAGCAAGAAAAGAGCTTATGTTGATATTTCTGAAAGA 1236
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1240 AAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGGTCTTTGGAGAGCTGTGAT 1299
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 1237 AATTTACTACGAGCGGTTGAGATGCAAGAGCTTTTGACTGGTATTTTGAAGAAACGTTTTT 1296
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1300 CCAGAACTGCTATGCTGTAGGACAGGCGCTTCTTTCGGGATGCTTCCAGAAACAAAGGA 1359
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 1297 CCTAATTTACATGTTCCAGAGGATAGACAGGCTGGCATGGGCTATTTCGAGTAGAGGG 1356
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1360 CTAAACAGACTACTGCTTTTGAATATAACCTTCCCGATGAAACACAGATTTGGGACACAG 1419
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 1357 ATCTCGTGAATGTTTAGATTTAATTTCTCTGACAAACAAACCCCAACAGGTGTAAACC-- 1414
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1420 GTCAATTTCTGTACCTGTGATGGATGGGCGGACAACTCAAGTCTTTCAGTACAGTCCAG 1479
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 1415 -TTTCACTGTTGGATGCCATGTGCAAGGAGGCAATCAATTTCTTGAATATCTTCAAC 1473
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
CC sample preferably serum or ovarian tissue. The method comprises
CC contacting a biological sample from a patient with (IV), detecting the
CC amount of polynucleotide hybridising to (IV) and comparing the amount to
CC a predetermined cutoff value and thereby detecting ovarian cancer in the
CC patient, where the amount of polynucleotide hybridising to (IV) is
CC detected preferably by polymerase chain reaction (PCR). (I) comprising
CC (iii) and/or (ii) is useful for stimulating and/or expanding T cells
CC specific for an ovarian tumour comprising contacting T cells
CC with (iii) or (ii). (iii) is useful in design and preparation of
CC ribozyme molecules for inhibiting expression of the tumour polypeptides
CC and proteins in tumour cells; and to isolate a full length gene from a
CC suitable library e.g., a tumour cDNA library using well known
CC techniques.

XX SQ Sequence 473 BP; 108 A; 164 C; 103 G; 98 T; 0 other;

Query Match 25.4%; Score 443.4; DB 24; Length 473;
Best Local Similarity 99.4%; Pred. No. 1.1e-81;
Matches 466; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 595 CTGATCCGCGCCCAACAGAGAGAGGGCTGTGGAGCCCGCTGCTGGGGCGCTCTGCG 654
DB 473 CTTATCCGCGCCCAACAGAGAGAGGGCTGTGGAGCCCGCTGCTGGGGCGCTCTGCG 414
QY 655 GCGAGGGCGGATGTTCTGACCTTCTGGACTGTCACTGTGAGTCCACAGAGGGTGGCTG 714
DB 413 GCGAGGGCGGATGTTCTGACCTTCTGGACTGTCACTGTGAGTCCACAGAGGGT-5CTG 355
QY 715 GAGCGCTGCTGCAGAGATCCATGAAGAGAGTCCGAGTGGTGTGCCCGGTGATTGAT 774
DB 354 GAGCGCTGCTGCAGAGATCCATGAAGAGAGTCCGAGTGGTGTGCCCGGTGATTGAT 295
QY 775 GTGATCGACTGGAAACCTTCCGAATACCTGGGAACTCCGGGAGCCCGAGATCGCGGT 834
DB 294 GTGATCGACTGGAAACCTTCCGAATACCTGGGAACTCCGGGAGCCCGAGATCGCGGT 235
QY 835 TTCGACTGGAGGCTGGTGTTCACG-TGGCACACAGTTCCTGAGAGGAGAGATACCGAT 893
DB 234 TTCGACTGGAGGCTGGTGTTCACGCTGGCACACAGTTCCTGAGAGGAGAGATACCGAT 175
QY 894 GCATCCCGCTGCATGCTCATCAGTCTCCACAAATGGCTGGTGGTGTGTTGCTGTGAG 953
DB 174 GCATCCCGCTGCATGCTCATCAGTCTCCACAAATGGCTGGTGGTGTGTTGCTGTGAG 115
QY 954 TAAGAAATATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGA 1013
DB 114 TAAGAAATATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGA 55
QY 1014 AAACCTCGAATTTTCCCTTTAGGATCTGGCAGTGTGGTGGGGTCTCGGAA 1062
DB 54 AAACCTCGAATTTTCCCTTTAGGATCTGGCAGTGTGGTGGGGTCTCGGAA 6

RESULT 15

ABL80705/c

ID ABL80705 standard; cDNA; 473 BP.

XX AC ABL80705;

XX DT 17-MAY-2002 (first entry)

XX DE Human ovarian cancer related cDNA clone SEQ ID NO:3683.

XX KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX OS Homo sapiens.

XX FN W0200192581-A2.

XX PD 06-DEC-2001.

XX XX

PF 29-MAY-2001; 2001WO-US17756.

XX PR 26-MAY-2000; 2000US-207484P.

XX PA (CORI-) CORIXA CORP.

XX PI Algate PA, Harlocker SL, Jones R;

XX DR WPI; 2002-122075/16.

XX PT Composition for therapy and diagnosis of ovarian cancer comprising

XX PT polypeptide of an ovarian tumor polypeptide, polynucleotide encoding

XX PT polypeptide, antibody specific to polypeptide or T cell expressing

XX PT polypeptide

XX PS Claim 1; SEQ ID 3683; 489pp; English.

XX CC The present invention describes a composition (I) comprising: carriers

XX CC and immunostimulants; and a polypeptide (II) of an ovarian tumour

XX CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence

XX CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to

XX CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell

XX CC population of (II), or antigen presenting cells that express (II).

XX CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to

XX CC sample preferably serum or ovarian tissue. The method comprises

XX CC contacting a biological sample from a patient with (IV), detecting the

XX CC amount of polynucleotide hybridising to (IV) and comparing the amount to

XX CC a predetermined cutoff value and thereby detecting ovarian cancer in the

XX CC patient, where the amount of polynucleotide hybridising to (IV) is

XX CC detected preferably by polymerase chain reaction (PCR). (I) comprising

XX CC (iii) and/or (ii) is useful for stimulating and/or expanding T cells

XX CC specific for an ovarian tumour protein comprising contacting T cells

XX CC with (iii) or (ii). (iii) is useful in design and preparation of

XX CC ribozyme molecules for inhibiting expression of the tumour polypeptides

XX CC and proteins in tumour cells; and to isolate a full length gene from a

XX CC suitable library e.g., a tumour cDNA library using well known

XX CC techniques.

XX SQ Sequence 473 BP; 106 A; 162 C; 105 G; 100 T; 0 other;

Query Match 24.4%; Score 425.8; DB 24; Length 473;

Best Local Similarity 98.7%; Pred. No. 4.5e-78;

Matches 471; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 592 CGCTGATCCGCCCAACAGAGAGAGGGCTGTGGAGCCCGCTGCTGGGGCGCTCT 651
DB 473 CGCTGATCCGCCCAACAGAGAGAGGGCTGTGGAGCCCGCTGCTGGGGCGCTCT 414
QY 652 GCGGCGAGGGCGGATGTTCTGACCTTCTGAGTGTCACTGTGAGTCCACGAGGGGTGG 711
DB 413 GCGGCGA-GGGCGATGTTCTGACCTTCTGAGTGTCACTGTGAGTCCACGAGGGT-G 356
QY 712 CTGGAGCCGCTGCTGCAGAGGATCCATGAAGAGAGTCCGACGTGTGTGTCGCCGGTATT 771
DB 355 CTGGAGCCGCTGCTGCAGAGGATCCATGAAGAGAGTCCGACGTGTGTGTCGCCGGTATT 296
QY 772 GATGATCGACTGGACACCTTCCGAATACCTCGGGNACTCCGGGGAGCCCGAGATCGGC 831
DB 295 GATGATCGACTGGACACCTTCCGAATACCTCGGGNACTCC-GGGAGCCCCAGATCGGC 237
QY 832 GGTTTCGACTGGAGGCTGGTGTTCAGTGTGCACACAGATTCTCTGAGAGGAGAGATACGG 891
DB 236 GGTTTCGACTGGAGGCTGGTGTTCAGTGTGCACACAGATTCTCTGAGAGGAGAGATACGG 177
QY 892 ATGCAATCCCGCTCGATGTCATCAGGTCTTCAACAAATGGCTGGTGGTGTGTTGCTGTG 951
DB 176 ATGCAATCCCGCTCGATGTCATCAGGTCTTCAACAAATGGCTGGTGGTGTGTTGCTGTG 117
QY 952 AGTAAGAAATATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGA 1011
DB 116 AGTAAGAAATATTTTGAATATCT-GGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGA 58

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 7, 2003, 12:45:48 ; Search time 6368.95 Seconds
(without alignments)
11215.070 Million cell updates/sec

Title: US-10-074-527-3
Perfect score: 1746
Sequence: 1 atgtgggggcacgcgcg.....tcaaagagcgtgtatga 1746

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 288711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_ats.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_ats.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1746	100.0	2745	6	AX477705 Sequence
2	1744.4	99.9	1746	9	AB078146 Homo sapi
3	1744.4	99.9	1746	9	AB078146 Homo sapi
4	1319	75.5	2343	9	AK024865 Homo sapi
5	851.4	48.8	6362	2	Continuation (5 of
6	851.4	48.8	110000	2	Continuation (3 of
7	851.4	48.8	288118	2	Continuation (8 of
8	842.6	48.3	228368	2	Continuation (8 of
9	841	48.2	94616	2	Continuation (8 of
10	841	48.2	288118	2	Continuation (8 of
11	830	47.5	1928	9	BC013945 Homo sapi
12	812.6	46.5	110000	2	Continuation (4 of
13	812.6	46.5	166547	2	Continuation (4 of
14	805.4	46.1	166547	2	Continuation (4 of
15	803.8	46.0	110000	2	Continuation (2 of
16	795.4	45.6	110000	2	Continuation (2 of
17	795.4	45.6	110000	2	Continuation (2 of
18	795.4	45.6	299134	2	Continuation (3 of
19	792.2	45.4	299134	2	Continuation (3 of
20	792.2	45.4	110000	2	Continuation (2 of
21	526.2	30.1	328901	2	Continuation (2 of
22	522.4	29.9	226623	2	Continuation (2 of
23	520.8	29.8	1737	10	U73819 Mus muscu
24	520.8	29.8	234109	2	Continuation (2 of
25	510.6	29.2	110000	2	Continuation (2 of
26	499.4	28.6	5369	9	BC036390 Homo sapi
27	499.4	28.6	73481	9	AC025034 Homo sapi
28	496.2	28.4	1737	9	AR236647 Sequence
29	496.2	28.4	1737	9	AR236647 Sequence
30	493.6	28.3	165175	2	Y08564 Homo sapien
31	419	24.0	554	6	AX197983 Sequence
32	418.4	24.0	544	6	AX208574 Sequence
33	400.8	23.0	652	6	AX197725 Sequence
34	400.2	22.9	632	6	AX208332 Sequence
35	400.2	22.9	632	6	AX209583 Sequence
36	374	21.4	149930	9	AL136084 Human DNA
37	363	20.8	483	6	AX371123 Sequence
38	300.8	17.2	2384	3	AK112431 Ciona int
39	285.6	16.4	1920	6	AR153422 Sequence
40	285.6	16.4	2575	6	AX376296 Sequence
41	285.6	16.4	2575	6	AX697278 Sequence
42	285.6	16.4	3357	9	AY035399 Homo sapi
43	285.6	16.4	3580	9	HS0803261
44	284	16.3	3332	6	AR155987 Sequence
45	284	16.3	3528	9	BC014789 Homo sapi

ALIGNMENTS

RESULT 1
AX477705
LOCUS AX477705
DEFINITION Sequence 17 from Patent WO0246426.
ACCESSION AX477705
VERSION AX477705.1
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Sanjanwala,M.M., Yao,M.G., Au-Young,J., Baughn,M.R., Arvizu,C.,
Ring,H.Z., Lee,E.A., Ding,L., Hafalia,A.J., Tang,Y.T., Yue,H.,
Tribouley,C.M., Lu,D.A., Warren,B.A., Yang,J.,

2745 bp DNA linear PAT 12-AUG-2002

[illegible]

961 TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAAACCTC 1020
Db TATTTTGAATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAAACCTC 1020
1021 GAATTTTCTTTAGGATCTGGCAGTGTGGGCTTCTGGAACACACCCATGTTCCCAT 1080
Db GAATTTTCTTTAGGATCTGGCAGTGTGGGCTTCTGGAACACACCCATGTTCCCAT 1080
1081 GTTGGCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGT 1140
Db GTTGGCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGT 1140
1141 GTTGGCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGT 1200
Db GTTGGCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGT 1200
1201 CGTGGCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGT 1260
Db CGTGGCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAGGCTCTGGCCAAACAGT 1260
1261 CAGGTAAAGACTTCAAGTGTCTTGGAGACTGTATCCAGAACTGCATGTCCTGAG 1320
Db CAGGTAAAGACTTCAAGTGTCTTGGAGACTGTATCCAGAACTGCATGTCCTGAG 1320
1321 GACAGGCTGCTTCTTGGGATGCTCCAGAAACAAAGGACTTAAACAGACTTCTGTTGAC 1380
Db GACAGGCTGCTTCTTGGGATGCTCCAGAAACAAAGGACTTAAACAGACTTCTGTTGAC 1380
1381 TATAACCTCCCGATGAACACAGATGTGGACACAGGCTCATCTGTACCTCTGTCAT 1440
Db TATAACCTCCCGATGAACACAGATGTGGACACAGGCTCATCTGTACCTCTGTCAT 1440
1441 GGGATGGCCAGAACTAGTCTTTCAGATACACGCTCCAGAAAGAAATACGCTATAACAC 1500
Db GGGATGGCCAGAACTAGTCTTTCAGATACACGCTCCAGAAAGAAATACGCTATAACAC 1500
1501 CACAGGCTGAGGCTGCTTCTTGGGATGCTCCAGAAACAAAGGACTTAAACAGACTTCTGTTGAC 1560
Db CACAGGCTGAGGCTGCTTCTTGGGATGCTCCAGAAACAAAGGACTTAAACAGACTTCTGTTGAC 1560
1561 TGGGAGAAATCTGCCAGAGAACTCAGATGTGGGACACAGGCTCATCTGTACCTCTGTCAT 1620
Db TGGGAGAAATCTGCCAGAGAACTCAGATGTGGGACACAGGCTCATCTGTACCTCTGTCAT 1620
1621 CACGAAACAGTCCAGAAATGTGTCCAGGCTCCAGAGGAGAGTGCAGTGCAGTTTCGTT 1680
Db CACGAAACAGTCCAGAAATGTGTCCAGGCTCCAGAGGAGAGTGCAGTGCAGTTTCGTT 1680
1681 CCACCTTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTTCAAAGAGCGCATG 1740
Db CCACCTTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTTCAAAGAGCGCATG 1740
1741 TTATGA 1746
Db TTATGA 1746

RESULT 4
AK024865
LOCUS AK024865 2343 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21212 fis, clone COL00502.
ACCESSION AK024865
VERSION AK024865.1 GI:10437273
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y.,
Okamoto S., Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T.,
Shibahara T., Tanaka T., Nakamura Y., Isogai T. and Sugano S.

TITLE
JOURNAL
REFERENCE
AUTHORS

NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2343)
Sugano S., Suzuki Y., Ota T., Obayashi M., Nishi T., Isogai T.,
Shibahara T., Tanaka T. and Nakamura Y.

TITLE
JOURNAL

Direct Submission
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail: cdnal@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,
Fax: 81-3-5449-5416)

COMMENT

NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- and 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES

source

Location/Qualifiers

1. 2343
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="COL00502"
/tissue_type="colon"
/clone_lib="COL"
/notes="cloning vector pME18SFL3"
465..1319
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB15027.1"
/db_xref="GI:10437274"

CDS

translation="MQSPVDVIRPTMAGLFAVSKYFEYLGSDYDTMBVWGENLE
FSFRIWQCGVLETHPCSHVGFVKQAPVSRNKALNSVRAAEVMDSEFKLYYHN
PRALPFGDVTERRKQLCKDEKWELETYYPELHPEDRPFGLMLQNKGLTDY
CFDYNPDENQIVGHQVILYLCHMGQNGFFEYTSQKEIRINTHQEGGCTAVKGMOT
LIMHLCEETAPENQKCFILOBDGSLFHEQSKKQVQARKSESSDFVPLLRDCTNSDHQK
WFFKRLM"

BASE COUNT 673 a 474 c 563 g 633 t
ORIGIN

Query Match

75.5%; Score 1319; DB 9; Length 2343;

Best Local Similarity 100.0%; Pred. No. 6.e-199;

Matches 1319; Conservaive 0; Mismatches 0; Indels 0; Gaps 0;

Qy 428 CATTTTATATGAAGCCTGTGCTCAACTCTCCTCGGACAGTTTACAGTGTCTTGAGACAT 487
Db 1 CATTTTATATGAAGCCTGTGCTCAACTCTCCTCGGACAGTTTACAGTGTCTTGAGACAT 60
Qy 488 CCCCGGATATCTCTGTAGAGAAAGTATCTCTGTAGATGACTACATGATAGAGAGCACC 547
Db 61 CCCCGGATATCTCTGTAGAGAAAGTATCTCTGTAGATGACTACATGATAGAGAGCACC 120
Qy 548 TGAAGGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCCAAGGTCGCGCTCATCGCGCCA 607
Db 121. TGAAGGAGCGCTTGGCCAAATGAGCTTTCGGGACTGCCCCAAGGTCGCGCTCATCGCGCCA 180
Qy 608 ACAAGAGAGAGCGCTTGGCCGAGCGCGCTGTGCGGGCGCTCTCGGCGAGGGCGCATG 667
Db 181 ACAAGAGAGAGCGCTTGGCCGAGCGCGCTGTGCGGGCGCTCTCGGCGAGGGCGCATG 240
Qy 668 TTCTGACCTTCTTGGACTGTCACTGTAGTGCCACGAAGGTCGCTGGAGCCGCTGCTGC 727
Db 241 TTCTGACCTTCTTGGACTGTCACTGTAGTGCCACGAAGGTCGCTGGAGCCGCTGCTGC 300
Qy 728 AGAGGATCCATGAAGAGGAGTGGCAGTGGTGTGCCCCGCTGATTGATGATGATGATGATG 787
Db 301 AGAGGATCCATGAAGAGGAGTGGCAGTGGTGTGCCCCGCTGATTGATGATGATGATGATG 360
Qy 788 ACACCTTCGAATACCTGGGGAACCTCGGGGAGCCCCAGATCGCGCGGTTTCGATCGAGGC 847
Db 361 ACACCTTCGAATACCTGGGGAACCTCGGGGAGCCCCAGATCGCGCGGTTTCGATCGAGGC 420

Qy 848 TGGTGTTCACGTGGCACACAGTTCCTCAGAGGGAGAGATACGATCAATCCCCCGTCG 907
Db |||||
Qy 421 TGGTGTTCACGTGGCACACAGTTCCTCAGAGGGAGAGATACGATCAATCCCCCGTCG 480
Db |||||
Qy 908 ATGTCATCAGGTCTCCCAACAATGGCTGGTGGCTGTTTGTGTGAGTAAGAAATATTTTG 967
Db |||||
Qy 481 ATGTCATCAGGTCTCCCAACAATGGCTGGTGGCTGTTTGTGTGAGTAAGAAATATTTTG 540
Db |||||
Qy 968 AATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAACCTCGAATTTT 1027
Db |||||
Qy 541 AATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAACCTCGAATTTT 600
Db |||||
Qy 1028 CCTTTAGGATCTGCAGTGGTGGGTCTCGAAACACACCCATGTTCCCATGTTGGCC 1087
Db |||||
Qy 601 CCTTTAGGATCTGCAGTGGTGGGTCTCGAAACACACCCATGTTCCCATGTTGGCC 660
Db |||||
Qy 1088 ATGTTTTCCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCCAACAGTGTTCGTG 1147
Db |||||
Qy 661 ATGTTTTCCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCCAACAGTGTTCGTG 720
Db |||||
Qy 1148 CAGCTGAAGTATGATGATGAATTTAAAGCTCTACTACCATCGCAACCCCGTGGCC 1207
Db |||||
Qy 721 CAGCTGAAGTATGATGATGAATTTAAAGCTCTACTACCATCGCAACCCCGTGGCC 780
Db |||||
Qy 1208 GCTTGGAACTTTTGGGGATGTGACAGAGGAAGAGCTCCCGGACAAAGCTCCAGTGTA 1267
Db |||||
Qy 781 GCTTGGAACTTTTGGGGATGTGACAGAGGAAGAGCTCCCGGACAAAGCTCCAGTGTA 840
Db |||||
Qy 1268 AAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGTCATGTGCTGAGGACAGGC 1327
Db |||||
Qy 841 AAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGTCATGTGCTGAGGACAGGC 900
Db |||||
Qy 1328 CTGCGTCTTCGGGATGCTCAGAAACAAAGACTTAACAGACTACTGCTTTGACTATTAAC 1387
Db |||||
Qy 901 CTGCGTCTTCGGGATGCTCAGAAACAAAGACTTAACAGACTACTGCTTTGACTATTAAC 960
Db |||||
Qy 1388 CTCCGATGAACCAAGTGTGGGACACAGGTCATTCTGCTCTGTCATGGGATGG 1447
Db |||||
Qy 961 CTCCGATGAACCAAGTGTGGGACACAGGTCATTCTGCTCTGTCATGGGATGG 1020
Db |||||
Qy 1448 GCCAGAATCAGTTTTTCAGTACACGTCCTCCAGAAAGAAATACGCTATAACACCCACAGC 1507
Db |||||
Qy 1021 GCCAGAATCAGTTTTTCAGTACACGTCCTCCAGAAAGAAATACGCTATAACACCCACAGC 1080
Db |||||
Qy 1508 CTGAGGCTGATGCTGTGGAAGCAGGAATGGAATACCTTATCATGATCTCTGCGAAG 1567
Db |||||
Qy 1081 CTGAGGCTGATGCTGTGGAAGCAGGAATGGAATACCTTATCATGATCTCTGCGAAG 1140
Db |||||
Qy 1568 AAAGTCCCGCAGAGATCAGAAATGATCTCTGCGAGGATGGATCTTTATTTCCAGAAC 1627
Db |||||
Qy 1141 AAAGTCCCGCAGAGATCAGAAATGATCTCTGCGAGGATGGATCTTTATTTCCAGAAC 1200
Db |||||
Qy 1628 AGTCCAGAAATGCTCCAGGCTGCGAGGAAGGAGTCCAGTGACAGTTCCTGTCACCTCT 1687
Db |||||
Qy 1201 AGTCCAGAAATGCTCCAGGCTGCGAGGAAGGAGTCCAGTGACAGTTCCTGTCACCTCT 1260
Db |||||
Qy 1688 TACGAGACTGACCAACTCGGATCATCAGAAATGTTCTTCAAAGAGGCGATGTTATGA 1746
Db |||||
Qy 1261 TACGAGACTGACCAACTCGGATCATCAGAAATGTTCTTCAAAGAGGCGATGTTATGA 1319
Db |||||

RESULT 5

AC132018_4

WPCOMMENT

Sequence split into 5 fragments LOCUS AC132018 Accession AC132018

Fragment Name	Begin	End
AC132018_0	1	110000
AC132018_1	100001	210000
AC132018_2	200001	310000
AC132018_3	300001	410000
AC132018_4	400001	463632

Continuation (5 of 5) of AC132018 from base 400001 (AC132018 Rattus norvegicus clone CH2

Query Match 48.8%; Score 851.4; DB 2; Length 63632;
Best Local Similarity 79.7%; Pred. No. 3.7e-125;
Matches 1081; Conservative 0; Mismatches 266; Indels 10; Gaps 6;
Qy 388 TATGATTATGATAAATTTGCCAGGACATCTGTTTATCATAGCATTTTATATGAAGCTGG 447
Db |||||
Qy 44126 TATGATTATGATAAATTTGCCAGGACATCCGTTGTCTATAGCATTTTATATGAAGCTGG 44185
Db |||||
Qy 448 TCACACTCTCTTCCGGACAGTTTACAGTCTCTTGGACATCCCGGATATCTCTGTAGAA 507
Db |||||
Qy 44186 TCACACTCTCTTCCGGACAGTTTACAGTCTCTTGGACATCCCGGATATCTCTGTAGAA 44245
Db |||||
Qy 508 GAAGTATCTCTTGTAGATGACTACAGTGTAGAGACACCTGAAGAGAGCCCTTGGCAAT 567
Db |||||
Qy 44246 GAGTCTATCTGTGTAGTACAGGACAGAGATCACCTGAAGAGAGCCCTTGGTCAA - 44304
Db |||||
Qy 568 GAGCTTTCCGGACCTGCCCCAAGGTGCGCTGTATCCCGGCCAACAGAGAGAGGCGCTGGTG 627
Db |||||
Qy 44305 -GAGCTTTCCAGCTGCGCTAGGTGCACTGTATCCCTGTAGCA - -AGAGAGGACCTAGTG 44361
Db |||||
Qy 628 CGAGCCCGGCTGCTGGGGGCTCTCGCGGAGGGGCGATGTTCTGACCTTCTCTGGACTGT 687
Db |||||
Qy 44362 CAAG -CGGGTACTTGGGAGCCTTTTGGGCGCAGGGGCGAAGTGTCTGAGGTTTCTTGGACTGT 44420
Db |||||
Qy 688 CACTGTGAGTCCCAAGGGTGGCTGGAGCCGCTGTGACAGAGGATCCCATGAAGAGGAG 747
Db |||||
Qy 44421 CACTGTGAGTCTCATGAGGGTGGCTGGAGTCCCTGTCTGAGAGGATCCACAGAGAGGAG 44480
Db |||||
Qy 748 TCGGCAGTGTGTGCCCGTGTGATGTGATCGATCGATGCAACACCTTTCGAATACCTGGGG 807
Db |||||
Qy 44481 TCGGCAGTGTGTGCCCGTGTGATGTGATCGATCGATGCAACACCTTTCGAATACCTGGGG 44540
Db |||||
Qy 808 AACTCCGGGAGGCCAGATCGCGGCTTTCGATCGAGGGCTGTGTTTACGTGCGACACA 867
Db |||||
Qy 44541 AACTCCGGGAGGCCAGATCGCGGCTTTCGATCGAGGGCTGTGTTTACGTGCGACAGTG 44600
Db |||||
Qy 868 GTTCTCTGAGAGGAGAGGATACCGATGCAATCCCGCTCGATGTCAATCAGGTCTTCAACA 927
Db |||||
Qy 44601 GTTCTCTGAGAGGAGAGGATACCGATGCAATCCCGCTCGATGTCAATCAGGTCTTCAACA 44660
Db |||||
Qy 928 ATGGCTGTGGGCTGTGTTGCTGTGAGTAAGAAATATTTTGAATATCTGGGGCTTATGAT 987
Db |||||
Qy 44661 ATGGCTGTGGGCTGTGTTGCTGTGAGTAAGAAATATTTTGAATATCTGGGGCTTATGAT 44720
Db |||||
Qy 988 ACAGGAATGCAATTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGSCAGTGT 1047
Db |||||
Qy 44721 ACCGGAATGGAAGTCTGGGAGGAGAAACCTTGAATTTCTCTTTAGGATCTGSCAGTGT 44780
Db |||||
Qy 1048 GGTGGGCTTCTGGAACACACACCATGTTCCCATGTTGGCCATGTTTTCCTCCCAAGCAAGCT 1107
Db |||||
Qy 44781 GGT - GACTCTGGAAACACACCATGTTCCCATGTTGGCCATGTTTTCCTCCCAAGCAAGCT 44838
Db |||||
Qy 1108 CCCTACTCCCGCAACAAAGGCTCTGGCCAAACAGTGTTCGTGAGTGAAGTATGATGGAT 1167
Db |||||
Qy 44839 CCCTACTCCCGCAACAAAGGCTCTGGCCAAACAGTGTTCGTGAGTGAAGTATGATGGAT 44898
Db |||||
Qy 1168 GAATTTAAAGAGTCTTACTACCATCGCAAACCCCGTCCCGCTTGGAACTTTTGGGGAT 1227
Db |||||
Qy 44899 GAATTTAAAGAGTCTTACTACCATCGCAAACCCCGTCCCGCTTGGAACTTTTGGGGAT 44957
Db |||||
Qy 1228 GTGACAGAGAGGAGAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGTGTTCTTG 1287
Db |||||
Qy 44958 GTGACAGAGAGAGAGAGCTTCCAGCTCCAGTGTGAGAGCTTAAAGCGGTTCTCTG 45017
Db |||||
Qy 1288 GAGACTGTATCCAGAACTGCAATGTGCTGAGGACAGGCTTGTCTTTCGGGATGCTC 1347
Db |||||
Qy 45018 GCTACTGTATCCAGAACTGCAATGTGCTGAGGACAGGCTTGTCTTTCGGGAAAGCTT 45077
Db |||||
Qy 1348 CAGAACAAAGGACTTAAACAGACTTCTGTTGACTATTAACCTCCCGATGAAACAGATT 1407
Db |||||
Qy 45078 GAGAACAGAGAGACTACCGGGGTACTGTCTGACTATTAATCTCTCCAGTGAANAATGTG 45137
Db |||||
Qy 1408 GTGGGACACCAAGGTCAATTCGTACCTCTGTCTGGGATGGGCCAGAAATCACTTTTTCGAG 1467
Db |||||

Db 45138 GAAGCCACACAGGTCCTTTTGTACTCTGCCACGGGATGGGTGAGAACCCAGTTTTCGAG 45197
Qy 1468 TACACGTCCTCCAGAAAGAAATACGCTATAACACCCACAGCCCTGAGGGCTGCATTGCTGTG 1527
Db 45198 TGTACATCCCGGCAAGAAATACGGTATAACACCCCGCCAGCCCGGAGTCTGATAGCAGTG 45257
Qy 1528 GAAGCAGGAATGGAATACCTTATCATGATCTCTGCGAAGAACTGCCGCCAGAGATCAG 1587
Db 45258 GAGGAAGGGAAGGAT--CCTTATCCCGGATCTCTGGAAGAGATACCGTTCCAGAGAAATCAA 45315
Qy 1588 AAGTTTCATCTGACAGGAGGATGATCTTTATTTTCCAGCAAGTCCCAAGAAATGTCAG 1647
Db 45316 GAGTTTCATCTACAGGAGGACGGACGTTAGTCTTACAGCAGACGGGAAATGTCGAG 45375
Qy 1648 GCTGCGAGGAAGGAGTCCGAGTGACAGTTTCGTTCCACTCTTACGAGACTGCACCAACTCG 1707
Db 45376 GCCACAGAGGAGGAGTTCAACTCTGCTCTGCTCCGAATTCGCGAGACTGTAGCAACTCA 45435
Qy 1708 GATCATCAGAAATGGTTCTTCAAGAGAGGAGGATGTTAT 1744
Db 45436 GACAGCCAGAGGTGGTCTCTTCAAGGAGACATGTCTAT 45472

RESULT 6

AC098559 2/c

WPCOMMENT

Sequence split into 8 fragments LOCUS AC098559 Accession AC098559

Fragment Name	Begin	End
AC098559_0	1	110000
AC098559_1	100001	210000
AC098559_2	200001	310000
AC098559_3	300001	410000
AC098559_4	400001	510000
AC098559_5	500001	610000
AC098559_6	600001	710000
AC098559_7	700001	794616

Continuation (3 of 8) of AC098559 from base 200001 (AC098559 Rattus norvegicus clone CH2

Query Match 48.8%; Score 851.4; DB 2; Length 110000;

Best Local Similarity 79.7%; Pred. No. 3.5e-125;

Matches 1081; Conservative 0; Mismatches 266; Indels 10; Gaps 6;

Qy 388 TATGATTATGATTAATTTGCCAGACATCTGTTATCATAGCATTTTATATGAAGCTCG 447
Db 95301 TATGATTATGATTAATTTGCCAGACATCTGTTATCATAGCATTTTATATGAAGCTCG 95242
Qy 448 TCACTCTCTCGGACAGTTTACAGTGTCTTGAGACATCCCGGATATCCTGCTAGAA 507
Db 95241 TCCACACTCTCTCAGACAGTTTACAGTGTCTTGAGACTTCCCTGTATATCCTGCTGGAG 95182
Qy 508 GAAGTGATCTCTGTAGATGACTACAGTGATAGAGACACCTGAAGGAGCGCTTGGCCAAAT 567
Db 95181 GAGGTCACTCTGTAGATGACTACAGCGACAGAGATCACCTGAAGGAGCGCTTGGTCAA - 95123
Qy 568 GAGCTTTCGGGATGTCCTCAAGGTGCGCTGATCCGCCCAACAGAGAGAGGCGCTGGTG 627
Db 95122 -GAGCTTTCACAGTGTCTGCTAGGTGCACTGATCCTCTGTAGCA - AGAGAGGACCTAGTG 95066
Qy 628 CGAGCCCGGCTGCTGGGGGCGTCTGCGGCGAGGGCGCATGTTCTGACCTTCTGGAAGTGT 687
Db 95065 CAAG -CGGGTACTGCGGAGCGCTTTCGCGCCAGGGGCGAAGTGTCTGAGGTTTCTGGAGTGT 95007
Qy 688 CACTGTGAGTCCACAGAGGAGTGTGCTGAGGCGGCTGCTGAGAGGATCCATGAAGAGGAG 747
Db 95006 CACTGTGAGTGTCTATGAGGGTGTGCTGAGTGTCTGCTGAGAGATCCACGAGAGAGAG 94947
Qy 748 TCGGCACTGTGTGCCCGGTGATTGATGATGATGATGATGATGATGATGATGATGATGATG 807
Db 94946 TCGGCGGTGTGTGCCCGGTGATTGATGATGATGATGATGATGATGATGATGATGATGATG 94887
Qy 808 AACTCCGGGAGGCCCAAGATCGGGCGGTTTGCATGAGAGGCTGTGTTTCACTGTCACACA 867

Db 94886 AACTCCGGAGAGCCCCAGATTGGTGGCTTTGACTGGCGCCTAGTATTATTCACGTGCGCAGTG 94827
Qy 868 GTTCTCTGAGAGGAGAGGATACCGGATGCAATCCCGCTCGATGTGTCATCAGGTCTCCAAACA 927
Db 94826 GTCCCTTGGCGGAGCGGAAGTTGATGCGGACTCTTATGACATTTATCAGGTCTCCAACT 94767
Qy 928 ATGCTGTGGTGGCTGTTTGTCTGTGAGTAAGAAATATTTGAATATCTGGGGTCTTATGAT 987
Db 94766 ATGCTGTGGGGAATGTTTGTGTTGAGTAAGAGATATTTGAATACCTGGGGTCTTATGAT 94707
Qy 988 ACAGGAATGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGCGAGTGT 1047
Db 94706 ACAGGAATGAAGTTTGGGAGGAGAAACCTTGAATTTCTTTAGGATCTGCGAGTGT 94647
Qy 1048 GGTGGGTTCTGGAACACACACCCATGTTCCCATGTTTGGCCATGTTTTCCTCCCAAGCAAGCT 1107
Db 94646 GGT - GACTCTGGAACACACACCCCTTACTCCCAAGTGGGCGACGCTTTCCTTAAGCAAGCT 94589
Qy 1108 CCCTACTCCGCAACAAAGGCTCTGGGCAACAGTGTTCGTGACGTGAAGTATGATGAT 1167
Db 94588 CCCTATTACGCGACAAAGGCTCTAGCCAAAGTGTCCGAGCTGCAGAGTGTGACGAGAT 94529
Qy 1168 GAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGGAACTTTTGGGGAT 1227
Db 94528 GAACTTAAAGACTCTACCAACCCCGGAAACCCCT - AGCCCGCTTGGAACTTTTGGGGAG 94470
Qy 1228 GTGACAGAGAGGAGAGCTCTCCGGGACAAAGCTCCAGTGTAAAGACTTTCAAGTGTCTTCTTG 1287
Db 94469 CTGACAGAGAGAAAGAGCTTTCAGCCCAAGCTCCAGTGTGAGACTTTAAAGCGGTCTCTG 94410
Qy 1288 GAGACTGTGTATCCAGAACTGCAATGTGCTGTGAGAGACAGGCTTCTTCTGGGATGCTC 1347
Db 94409 GCTACACTGTATCCAGAACTGTCATGTGCCAGAGACAGGCTGCTTCTTCTGGGAAAGCTT 94350
Qy 1348 CAGAACAAAGGACTTAAAGACTACTCTGTTGACTATTAACCTCCCGATGAAACCAAGATT 1407
Db 94349 GAGAACAGAGACTTACGGGGTACTGCTGGACTATTAATCTCCAGTGAAACCAATGTG 94290
Qy 1408 GTGGGACACAGAGTCAATCTGTACCTCTGTATGGGATGGCGGAGATCAGTTTTCGAG 1467
Db 94289 GAAGGCCACAGGCTCTTTTGTACTCTGCGACGGGATGGGTGAGAACAGTCTTTCGAG 94230
Qy 1468 TACAGCTCCAGAAAGAAATACGCTATAACCCCAACAGCTGAGGCTGAGGCTGATTTGCTGTG 1527
Db 94229 TGTACATCCCGGCAAGAAATACGGTATAACCCCGCAGGCCAGGCTGTCATAGCAGTG 94170
Qy 1528 GAAGCAGGATGATACCTTATCATGCACTCTGCGAAGAACTGCCCGAGAGAAATCAG 1587
Db 94169 GAGGAAGGGAAGGAT - CCTTATCCCGGATCTCTGGAAGATACCCGTTCCAGAGAAATCAA 94112
Qy 1588 AAGTTTCATCTTGCAGGAGGATGATCTTTATTTTCAAGACAGTCCCAAGAAATGTGTCAG 1647
Db 94111 GAGTTTCATCTTACAGGAGGACGCGCTAGTCTACAGCAGAGCGGGAATGTGTCAG 94052
Qy 1648 GCTGCGAGGAAGGAGTGCAGTGCAGTGTTCGTTTCCACTCTTACGAGACTGCACCAACTCG 1707
Db 94051 GCCACAGAGGAGGAGTTCAACTCTGCTTGTCTCCGAACTTGGCAGACTGTAGCAACTCA 93992
Qy 1708 GATCATCAGAAATGTTTCTTCAAGAGCGCATCTTAT 1744
Db 93991 GACAGCCAGAGGTGGTCTTCAAGGAGACATGTCTAT 93955

RESULT 7

AC129139

LOCUS

DEFINITION

Rattus norvegicus clone CH230-293111, *** SEQUENCING IN PROGRESS

AC129139

VERSION

AC129139.2 GI:23829088

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus

AC129139 288118 bp DNA linear HTG 11-OCT-2002
Rattus norvegicus clone CH230-293111, *** SEQUENCING IN PROGRESS
***, 14 unordered pieces.
AC129139
AC129139.2 GI:23829088
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 288118)
 Muzny,D,Marie, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsebrook, S., Anin, A., Anguiano, D., Anyalebechi, V., Ayoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Blawie, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Haviak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Louised, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, D., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sander, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Snajds, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejdos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

TITLE

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (27-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

TITLE

JOURNAL

3 (bases 1 to 288118)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Oct 11, 2002 this sequence version replaced gi:21989925. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Aclis (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GRGM

Center clone name: CH230-293111

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 214989 bases at least Q40

Consensus quality: 22817 bases at least Q30

Consensus quality: 27320 bases at least Q20

Estimated insert size: 252471; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)

* NOTE: This sequence may represent more than one clone.

* NOTE: This is a 'working draft' sequence. It currently

consists of 14 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1	5778:	contig of 5778 bp in length
*	5878:	gap of unknown length
*	5879	16373: contig of 10495 bp in length
*	16374:	gap of unknown length
*	16474	90430: contig of 73947 bp in length
*	90421	90520: gap of unknown length
*	90521	236782: contig of 146262 bp in length
*	236783	236882: gap of unknown length
*	236883	250122: contig of 13240 bp in length
*	250123	250222: gap of unknown length
*	250223	257033: contig of 6811 bp in length
*	257034	257133: gap of unknown length
*	259071	259071: contig of 1938 bp in length
*	259072	259171: gap of unknown length
*	259172	261008: contig of 1837 bp in length
*	261009	261108: gap of unknown length
*	264448	contig of 3340 bp in length
*	264449	gap of unknown length
*	264549	contig of 3000 bp in length
*	267549	267548: gap of unknown length
*	267649	271482: contig of 3834 bp in length
*	271483	271582: gap of unknown length
*	271583	275622: contig of 4040 bp in length
*	275623	275722: gap of unknown length
*	275723	286107: contig of 10385 bp in length
*	286108	286207: gap of unknown length
*	286208	288118: contig of 1911 bp in length.

FEATURES
 source

1..288118
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-293111"

1..1287

/note="wgs_end_extension"

clone_end:T7"

misc_feature

1338..5778

/note="wgs_end_extension"

clone_end:T7"

misc_feature	3086..3988 /note="clone_boundary clone_end:T7 site:MboI end_sequence:RXAIQ54TJB" 5879..7959 /note="wgs contig" 12280..13622 /note="wgs contig" 14019..16373 /note="wgs contig" 16474..19067 /note="wgs contig" 19118..21281 /note="wgs contig" 21332..22964 /note="wgs contig" 28345..31644 /note="wgs contig" 31695..34111 /note="wgs contig" 40505..44913 /note="wgs contig" 44964..47869 /note="wgs contig" 47920..49820 /note="wgs contig" 49871..52672 /note="wgs contig" 52723..56843 /note="wgs contig" 61887..63717 /note="wgs contig" 72165..76565 /note="wgs contig" 78769..86319 /note="wgs contig" 86370..90420 /note="wgs contig" 90521..92438 /note="wgs contig" 92489..94959 /note="wgs contig" 96097..99542 /note="wgs contig" 236883..239542	83876 CACTGTGACTGTCTCATGAGGGGTGGCTGGAGTCCCTGCTGCAGAGGATCCACGAGGAG 83933
misc_feature	48.8%; Score 851.4; DB 2; Length 288118; Best Local Similarity 79.1%; Pred. No. 3.1e-125; Matches 1081; Conservative 0; Mismatches 266; Indels 10; Gaps 6;	748 TCGGCAGTGGTGTGCCCGGTGATTGATGCAGCTGGAAACACCTTCGATAATCCTCGGG 807
misc_feature	388 TATGATTATGATAAATTGCCAGGACATCTGTTATCATAGCATTTTATATCAAGCCTGG 447	83936 TCGGCGGTGTGTGCCCGGTATTCGATGTCATTGACCTGGAAACACCTCTGAGTACCTCGGC 83995
misc_feature	83581 TATGATTATGATAACTCTGCCAAGACATCCGTTGTCTAGACATTTTATTAAGAAGCCTGG 83640	808 AACTCCGGGGAGCCCCAGATCGGCGGTTTCGACTGGAGGCTGTGTTTCAGTGCAGCACA 867
misc_feature	448 TCAACTCTCTTGGGACAGTTTACAGTGTCTTGGAGACATCCCGGATATCTGCTAGAA 507	83996 AACTCCGAGAGCCCCAGATTGGTGGCTTTGACTGGCGCTAGTATTTACGTCGACAGTG 84055
misc_feature	83641 TCCACACTCTTCAGACAGTTTACAGTGTCTCGAGCTTCCCTGATATCTGCTGGAG 83700	868 GTTCCTGAGAGGAGAGATACGATCAATCCCCCGTCAATGTCATGTCATCAGGTCTCCACA 927
misc_feature	508 GAAGTGATCTTGTAGATGACTACAGTGATAGAGACACTGAAAGAGCGCTTGCCCAAT 567	84056 GTCCCTTGGCGGAGCGGAAGTTGATCGGACTCTATTGACATATTACGCTCTCCAACT 84115
misc_feature	83701 GAGGTCAATCTGGTAGATGACTACAGCGACAGAGATCACTGAAAGAGCCCCCTGCTCAA- 83759	928 ATGGCTGTGGGCTGTTTGTCTGTGAGTAAGAAATATTTTCAATATCTGGGGTCTTATGAT 987
misc_feature	568 GAGCTTTCGGGACTGCCAAGGTGGCTGATCCCGCCCAACAGAGAGAGGCGCTGCTG 627	84116 ATGGCTGGGGACTGTTTGTGTGAGTAAAGATATTTTGAATACCTGGGGTCTTATGAT 84175
misc_feature	83760 -GAGCTTCAGCTGCCCTAGGTGACACCTGATCCCTGCTAGCA--AGAGAGGACCTAGTG 83816	988 ACAGGAATGGAAGTTTGGGAGGAGAAAACTCGAAATTTTCTTTAGGATCTGGCAGTGT 1047
misc_feature	628 CGAGCCCGGCTGCTGGGGCGCTGTCGGCGAGGGCGATGTTCTGACCTTCTCGACTGT 687	84176 ACCGGAATGGAAGTCTGGGAGGAGAAAACTTGAATTTCTCTTTAGGATCTGGCAGTGT 84235
misc_feature	83817 CAAG-CCGCTACTGGAGCCCTTGGGCGCAGGGCGAGTCTGAGTTTCTGACTGT 83875	1048 GGTGGGTTCTGGNAACACACCCCATGTTCCCATGTTGGCCATGTTTTCCTCCCAAGCAAGCT 1107
misc_feature	688 CACTGTGAGTGCCAGAAAGGTGGCTGGAGCGGCTGCTGCAGAGGATCCATGAAGAGGAG 747	84236 GGT--GACTCTGGAAACACACCCCTTACTCCACGTGGGCCACGCTTCTCCCTAAAGCAAGCT 84293
		1108 CCTACTCCGCAACAAGGCTCTGGCCAAACAGTGTTCGTGCAGCTGAAGTATCGATGGAT 1167
		84294 CCTATTCA GCGACGAAGGCTCTAGCCAAACAGTGTCCGACTGCGAAGTGTGGACGAT 84353
		1168 GAATTTAAAGAGCTCTACTACCATCGCAACCCCGCTGCCCGCTTGGAAACCTTTTGGGGAT 1227
		84354 GAACCTTAAAGACTCTTACCAACCCCGCAACCCCC-AGCCCGCTTGGAAACCTCTGGGGAG 84412
		1228 GTGACAGAGAGGAAGCAGCTCCGGGCAACAGCTCAAGTGTAAAGACTTCAAGTGGTCTTG 1287
		84413 CTGACAGAGAGAAAGAGCTTTCAGCCCAAGCTCCAGCTGTGAAGACTTAAAGCGGTTCCCTG 84472
		1288 GAGACTGTGTATCCAGAACTGCAATGTCCTGAGGACAGGCGCTCGCTTCTTCGGGATGCTC 1347
		84473 GCTACACTGTACCAGAACTGTCATGTGCCAGAGCAAGGCTCGTTCCTTCGGGAAGCTT 84532
		1348 CAGAACAAAGGACTTACAGACTACTGCTTTGACTATAAACCTCCCGATGAAACCAAGATT 1407
		84533 GAGAACAGAGACTTACGGGGGTACTGCTGACTATAAATCTCCAGTGAACAATGTG 84592
		1408 GTGGGACACAGGTCAATCTGTACCTCTGTATGGAGTGGCCAGAAATCAAGTTTTCGAG 1467
		84593 GAAGGCCACCAAGTCTCTTTGTACCTCTGCAAGGATGGGTGAGAACCAAGTTTTCGAG 84652
		1468 TACAGTCTCCAGAGAAATACGCTATACACCCACCGCTGAGGGCTGCATTCGCTGTG 1527
		84653 TGTACATCCCGCAAGAAATACGGGTATAACCCCGCAGCCCGGCTGTCATAGCAGTG 84712
		1528 GAAGCAGGAATGGATACCTTTATCTGCTATGCTCTCTGGGAAGAAACTGCCCCAGAGAATCAG 1587
		84713 GAGGAAGGGAAGGAT--CCTTATCCCGGATCTCTGGAAGATACCGTTCAGAGAGATCAA 84770
		1588 AGTTTCATCTTGCAGGAGGATGATCTTTATTTTCA GAAACAGTCCAAAGAAATGTTGCCAG 1647
		84771 GAGTTTCACTCTACAGGAGGACGGCAGCTTAGTCTACAAGCAGAGCGGGAATGTTGGAG 84830
		1648 GCTCGAGGAGGAGTCCGATGACAGTTTGGTTCGCTCTTACGAGACTGCACCAACTCG 1707
		84831 GCCACAGAGAGGAGTTTCAACTCTGCTTGTCTCCGAACCTTGCAGACTGTAGCAACTCA 84890
		1708 GATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTAT 1744
		84891 GACAGCCNAGAGTGGTCTTCAAGGAGCACATGTCAT 84927
		RESULT 8 AC121727

LOCUS AC121727 228368 bp DNA linear HTG 26-SEP-2002
 DEFINITION Rattus norvegicus clone CH230-326G7, *** SEQUENCING IN PROGRESS
 *, 13 unordered pieces.
 AC121727
 AC121727.3 GI:23322328
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 Rattus.
 1 (bases 1 to 228368)
 Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyie, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gbrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenauhewa, L., Loulseged, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapus, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okunju, G., Olarnpunsagoon, A., Pal, S., Parks, K., Paternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poldexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Speed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.
 Direct Submission
 Unpublished
 2 (bases 1 to 228368)
 Worley, K. C.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (21-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 228368)
 Rat Genome Sequencing Consortium.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (26-SEP-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Sep 26, 2002 this sequence version replaced gi:21909383.
 The sequence in this assembly is a combination of BAC-based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
 ----- Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GYGF
 Center clone name: CH230-326G7
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 174976 bases at least Q40
 Consensus quality: 181861 bases at least Q30
 Consensus quality: 185371 bases at least Q20
 Estimated insert size: 239944; sum-of-contigs estimation
 Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 3598: contig of 3598 bp in length
 * 3599 3698: gap of unknown length
 * 3999 14450: contig of 10752 bp in length
 * 14551 14550: gap of unknown length
 * 14551 49626: contig of 34976 bp in length
 * 49527 49626: gap of unknown length
 * 49627 59763: contig of 10137 bp in length
 * 59764 59863: gap of unknown length
 * 59864 89249: contig of 29386 bp in length
 * 89250 89349: gap of unknown length
 * 89350 95260: contig of 5911 bp in length
 * 95261 95360: gap of unknown length
 * 95361 202485: contig of 107125 bp in length
 * 202486 202585: gap of unknown length
 * 202586 214412: contig of 11827 bp in length
 * 214413 214513: gap of unknown length
 * 214513 216209: contig of 1697 bp in length
 * 216210 216309: gap of unknown length
 * 216310 217761: contig of 1452 bp in length
 * 217662 217861: gap of unknown length
 * 217862 219647: contig of 1786 bp in length
 * 219648 219748: gap of unknown length
 * 219748 224650: contig of 4902 bp in length
 * 224650 224749: gap of unknown length
 * 224750 228368: contig of 3619 bp in length.

FEATURES
 Location/Qualifiers
 1. 228368
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-326G7"
 14551..17292
 /note="wgs contig"
 22272..24156
 /note="wgs contig"
 40041..43821
 /note="wgs contig"

misc_feature 49627..51977
/note="wgs contig"
misc_feature 54058..59763
/note="wgs contig"
misc_feature 89350..90810
/note="wgs contig"
misc_feature 90861..93339
/note="wgs contig"
misc_feature 93390..95260
/note="wgs contig"
misc_feature 95361..97696
/note="wgs contig"
misc_feature 192296..194665
/note="wgs contig"
misc_feature 202586..204894
/note="wgs contig"
misc_feature 212212..214412
/note="wgs contig"
BASE COUNT 54283 a 40893 c 40157 g 51976 t 41059 others
ORIGIN

Query Match 48.3%; Score 842.6; DB 2; Length 228368;
Best Local Similarity 79.6%; Pred. No. 7.8e-124;
Matches 1084; Conservative 0; Mismatches 264; Indels 13; Gaps 7;
Qy 388 TATGATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 447
Db 163632 TATGATTATGATAACCTGCCCAAGACATCCGTTGTCATAGCATTTTATATGAAGCCTGG 163691
Qy 448 TCAACTCTCTCCGAGCAGTTTACAGTGTCTTCCAGACATCCCGGATATCTCTAGAA 507
Db 163692 TCCACATCTCTCCGAGCAGTTTACAGTGTCTTCCAGACATCTCCCGTATATCTCTGGAG 163751
Qy 508 GAAGTGATCTCTGTAGATGACTACAGTGATAGAGACACCTGAAGGAGCGCTTGGCCAAT 567
Db 163752 GAGTCAATTTGGTAGATGACTACAGTGACAGAGATCACCTGAAGGAGCGCTTGGCCA - 163810
Qy 568 GAGCTTTCCGAGCTGCCCAAGGTGGCTGATCGCGGCCAA CAAGAGAGAGGGCTGGTG 627
Db 163811 GAGCTGTACAGCTGCCCTAGGTGACCTGATCCCTGTAGCA - AGAGAGGGCTTAGTG 163868
Qy 628 CGAGCCGGCTGCTGGGGGCTGCTGGCGGAGGGGCGATGTTCTGACCTTCTGGAGCTGT 687
Db 163869 CAAG - CCGGCTACTGGGAGGCTTTGGCGGCCAGGGGCGAAGTGTCTGGAGCTGT 163927
Qy 688 CACTGTGAGTGCCACGAAGGCTGGCTGGAGCGCTGTGACAGAGGATCCATGAAGAGGAG 747
Db 163928 CACTGTGAGTGTATGAGGGGTGGCTGGAGTCCCTGTGACAGAGATCCACGAGAGGAG 163987
Qy 748 TCGCAGTGTGTGCCCGGTGATTTGATGTATGATCGACTGGAAACACCTTCGAATACCTGGGG 807
Db 163988 TCGCCCGTGTGTGCCCGGTATCGATGTATTGACTGGAAACACCTCTGAGTACCTGGGC 164047
Qy 808 AACTCCGGGAGGCCAGATCGGCGGTTTCACTGGAGGCTGGTTTCACTGGGACACACA 867
Db 164048 AACTCCAGAGAGGCCAGATGGTGGCTTTGACTGGCGCTTAGTATTTTCACTGGGACAGTG 164107
Qy 868 GTTCTCGAGAGGAGAGATACGATGCAATCCCGCTGATGTCATCAGTGTCTCCAAACA 927
Db 164108 GTCCCTTGGGGAGCGGAGTTGATCGGAGTCTTATGACATATCAGTGTCTCCAACT 164167
Qy 928 ATGGCTGGTGGCTGTTTGTGTGAGTAAGAAATATTTTGAATATCTGGGCTCTTATGAT 987
Db 164168 ATGGCTGGGAGCTGTTTGTGTGAGTAAGAGATATTTTGAATATCTGGGCTCTTATGAT 164227
Qy 988 ACAGGAATGAGTTTGGGAGGAGAAACCTCGAATTTTCTTTTAGATCTGGCAGTGT 1047
Db 164228 ACCGGAATGAGTGTCTGGGAGGAGAAACCTTGAATTTCTCTTTTAGATCTGGCAGTGT 164287
Qy 1048 GGTGGGTTCTGGAACACACACCAATGTTCCCATGTTGGCCATGTTTCCCGCAGCAAGCT 1107
Db 164288 GGT -- GACTCTGGAACACACACCTTACTCCCACTGGGGCAGCTCTTCCCTTAAGCAAGCT 164345

Qy 1108 CCTACTCCCGCAACAAGGCTCTGGCCAA CAGTGTTCGTGAGCTGAAGTATGATGGAT 1167
Db 164346 CCTATTTCAGCGACGAAGGCTCTAGCCAA CAGTGTTCGTGAGCTGAAGTATGATGGAT 164405
Qy 1168 GAATTTAAGAGCTCTACTACCATCGCAACCCCGTGGCTTGGAACTTTTGGGGAT 1227
Db 164406 GAATTTAAGAACTCTACTACCCCGGAAACCCCG - AGCCCGCTGGAACTTCTGGGGAG 164464
Qy 1228 GTGACAGAGAGGAGAGCTCCCGGCAAGCTCCAGTGTAAAGACTTCAAGTGTCTTTG 1287
Db 164465 CTGACAGAGAGAGAGCTTCCAGCCAGCTCCAGTGTGAAGACTTANAGCGGTTCTTG 164524
Qy 1288 GAGACTGTGTATCCAGAACTGTCATGTGCTGAGGAGGCTGCTTCTTGGGATGCTC 1347
Db 164525 GCTACAAATGTACCCAGAACTGTCATGTGCCAGAGGACAGGCTGCTTCTTCCGGAAGCTT 164584
Qy 1348 CAGAACAAAGGACTTACAGACTACTGCTTTTGTACTATTAACCTCCCGATGAAACCAAGATT 1407
Db 164585 GAGAACAGAGGAGACGCGGGGTACTGCTGGATATTAATCTCCCACTGAACCAAGTGTG 164644
Qy 1408 GTGGGACACCAAGGCTCATTTCTGTACTCTGTATGGGATGGGCCAGAAATCAGTTTTTCGAG 1467
Db 164645 GAAGGCCACCAAGGCTCTTTTGTACTCTGTCCACGAGATGGGTGAGAACCAAGTTTTCGAG 164704
Qy 1468 TACAGCTCCAGAGAAAGAAATACGCTATAACACCAACAGCTGAGGCTGAGGCTGCTGTGTG 1527
Db 164705 TGTACATCCCGGCAAGAAATACGCTATAACACCCGACGCCAGGCTGCTGATAGCAGTG 164764
Qy 1528 GAAGCAGGATGATACCTTTATCATGCTATCTGCGAAGAACTGCCCGCAGAAATCAG 1587
Db 164765 GAGAACGGAAGGAT -- CCTTATTCGGGATCTCTGGAAGATACGGTTCACAGAAATCAA 164822
Qy 1588 AAGTTCACTTTCAGGAGGATGGA --- TCTTTATTTACGAAACAGTCCAAGAAATGTGT 1643
Db 164823 GAGTTCACTTTCAGGAGGAGCAGCAGCAGTGTAGTCTACAGCAGACCGGAAATGTGT 164882
Qy 1644 CCAGGCTGCGAAGAGAGTGCAGTGCAGTGTTCGTTTCCACTCTTACGAGACTGCACCAA 1703
Db 164883 GGAGGCCACAGAGAGGAGTTCAACTCTGCTTTTGTCTCCGAACTTTCGAGACTGTAGCAA 164942
Qy 1704 CTGGATCATCAGAAATGGTTCTTCAAGAGCGCATGTTAT 1744
Db 164943 CTCAGACGCCAGGAGTGTCTGCCAGGAGCTCATGTCTAT 164983

RESULT 9
AC098559 7/c
WPCOMMENT

Sequence split into 8 fragments LOCUS AC098559 Accession AC098559

Fragment Name	Begin	End
AC098559_0	1	110000
AC098559_1	100001	210000
AC098559_2	200001	310000
AC098559_3	300001	410000
AC098559_4	400001	510000
AC098559_5	500001	610000
AC098559_6	600001	710000
AC098559_7	700001	794616

Continuation (8 of 8) of AC098559 from Base 700001 (AC098559 Ratus norvegicus clone CH2)

Query Match 48.2%; Score 841; DB 2; Length 94616;
Best Local Similarity 79.6%; Pred. No. 1.5e-123;
Matches 1083; Conservative 0; Mismatches 265; Indels 13; Gaps 7;

Qy 388 TATGATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 447
Db 61744 TATGATTATGATAACCTGCCCAAGACATCCGTTGTCTATAGCATTTTATATGAAGCCTGG 61685
Qy 448 TCAACTCTCTCCGAGCAGTTTACAGTGTCTTTCAGACATCCCGGATATCTCTGCTAGAA 507
Db 61684 TCCACACTCTCTCAGACAGTTTACAGTGTCTCAGAGATTTCCCGCTGATATCTCTGCTGAG 61625
Qy 508 GAAGTGATCTCTGTAGATGACTACAGTGATAGAGACCTGAAGGAGCGCTTGGCCAAT 567

Db 61624 GAGGTCATTTTGGTAGTAGTACAGTACAGAGATCACCTGAAGAGAGCCCTCGCCAA- 61566
Qy 568 GAGCTTTCGAGACTGCCCCAGGTCGGCTGATCGGCCCAACAGAGAGAGGGGCTCGTG 627
Db 61565 GAGCTGTACAGAGTGCCTAGGTGACCTGATCCCTGCTAGCA--AGAGAGGGGCTTAGTG 61508
Qy 628 CGAGCCCGCTGCTGGGGGCGCTGTCGGCGAGGGGCGATGTTCTGACCTTCTCGACTGT 687
Db 61507 CAAG-CCGGCTACTGGGAGCCCTTTGGCGCCAGGGGCGAGTGTCTGAGGTTCTCGACTGT 61449
Qy 688 CACTGTGAGTCCACGAGAGGCTGGAGCGCTGCTGCAGAGGATCCATGAAGAGAG 747
Db 61448 CACTGTGACTGTATAGGGGCTGGCTGGAGTCCCTGCTGCAGAGGATCCACGAGAGGAG 61389
Qy 748 TCGGAGTGGTGTGCGCGGTATGATGTGATCACTGGACACACCTTCGATACCTCGGG 807
Db 61388 TCGGCGGTGGTGTGCGCGGTATGATGTGATCACTGGACACACCTTCGATACCTCGGG 61329
Qy 808 AACTCCGGGAGCCCCAGATCGGGGTTTCACTGGAGGCTGGTTCACGTGCACACACA 867
Db 61328 AACTCCAGAGAGCCCCAGATGGTGGCTTTCGACTGGCGCTAGTATTCACGTGCACAGTG 61269
Qy 868 GTTCTCTGAGAGGAGAGATCGGATGCAATCCCGCTCGATGTATCAGGTCTCCAAACA 927
Db 61268 GTCCCTTGGCGGAGCGGAAGTTGATGCGGACTCCTATTGACATTATCAGGTCTCCAACT 61209
Qy 928 ATGCTGTGGCTGTGTTGCTGTGATGAAGAAATATTTGATATCTGGGTCTTATGAT 987
Db 61208 ATGCTGTGGGAGCTGTTGTTGTGTGATGAAGATATTTTGAATACCTGGGTCTTATGAT 61149
Qy 988 ACAGAAATGGAAGTTTGGGAGGAGAAACCTCGAATTTCTCTTAGGATCTGGCAGTGT 1047
Db 61148 ACCGAATGGAAGTCTGGGAGGAGAAACCTTGAATTTCTCTTAGGATCTGGCAGTGT 61089
Qy 1048 GGTGGGTTCTGGAACACACCCATGTTTCCCATGTTGGCCATGTTTTCCTCCCAAGCAAGCT 1107
Db 61088 GGT--GACTCAGGAACACACCCCTTACTCCCAAGTGGGCCATTTCTTCCCTTAAGCAAGCT 61031
Qy 1108 CCCTACTCCCGAACAGGCTCTGCCAACAGTGTTCGTGACGTGAAGTATGATGAT 1167
Db 61030 CCCTATTACGACGACGAGGCTCTAGCCAACAGTGTCCGAGCTGCAGAGTGTGACCGAT 60971
Qy 1168 GAATTTAAGAGCTTACTACATCGCAACCCCGTGCCTGCGGCTTGGAACTTTTGGGAT 1227
Db 60970 GAATTTAAGAGCTTACTACATCGCAACCCCGGAAACCCCG--AGCCCGCTTGGAACTTCTGGGAG 60912
Qy 1228 GTGACAGAGAGGAGCAGCTCGGGACAAAGCTTCCAGTGTAAAGACTTCAAGTGGTCTTG 1287
Db 60911 CTGACAGAGAGAGAGAGCTTCAGCCCAAGCTCCAGTGTGAAGACTTAAAGCGGTTCCTG 60852
Qy 1288 GAGACTGTGTATCGAAGACTGTGCTGAGGACAGGCTGCTTCTCGGATGCTC 1347
Db 60851 GCTCAATTTTACCCAGAACTGATGTGCGCAGAGGACAGGCTGTTCTTCGGGAAGCTT 60792
Qy 1348 CAGAAACAGAGCTAACAGACTACTGCTTTCAGTATTAACCTCCCGATGAACCAAGATT 1407
Db 60791 GAGAAACAGAGAGAGGAGGCTGCTGCTGCTGACTATATCTCCAGTGAACCAAGTGTG 60732
Qy 1408 GTGGACACCAAGGCTTCTGTACCTCTGTATGGGATGGGCCAGAAATCAGTTTTTCGAG 1467
Db 60731 GAAGGCCACCAAGTCTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60672
Qy 1468 TACAGTCTCCAGAAAGATACGCTATTAACACCCACAGCTGAGGCTGCTGCTGCTGCTG 1527
Db 60671 TGTACATCCCGCAGAAATACGCTATTAACACCCCGCAGGCTGCTGCTGCTGCTGCTG 60612
Qy 1528 GAAGCAGAAATGATACCTTATCATGCTCTCTCGAAGAACTTCCGCCAGAGAAATCAG 1587
Db 60611 GAGNAGGAGAGGAT--CCTATTCCGATCTCTGGAAGATACGGTTCGAGAGATCAA 60554
Qy 1588 AAGTTTCATCTTCAGGAGGATGGA-----TCTTTATTTCCAGAACAGTCCAGAAATGTT 1643

Db 60553 GAGTTCATCTACAGGAGGAGCGACAGCAGCTTAGTCTACAGCAGACCGGAAATGTGT 60494
Qy 1644 CCAGGCTGCGAGGAGGAGTCGAGTGCACAGCTTTCGTTCCACTCTTACGAGACTGCACAA 1703
Db 60493 GAGGCCACAGAGAGGAGTTCACCTGCTTTCGTTCCGAACTTGGCAGACTGTAGCAA 60434
Qy 1704 CTCGCTATCATFAGAAATGTTTCTTCAAAGAGCCCATGTTAT 1744
Db 60433 CTCAGACAGCAGAGGTTCTTCTGCCAGGAGCTCATGTCTAT 60393

RESULT 10

AC129139/c
LOCUS

DEFINITION

AC129139

HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.

Rattus norvegicus (Norway rat)

ORGANISM

REFERENCE

AUTHORS

AC129139 288118 bp DNA linear HTG 11-OCT-2002
Rattus norvegicus clone CH230-293111, *** SEQUENCING IN PROGRESS
***, 14 unordered pieces.
AC129139 GI:23829088
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 288118)
Mullen, D., Marie, H., Metzger, M., Lee, S., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrechts, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S., L. Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovat, C.,
Kowis, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Loulseg, H., Lozano, R., Lu, X., Lu, X., Ma, J.,
Maheshwari, M., Mahdarte, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mathew, S., McLeod, M., McNeill, T., McNeill, T., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokeleneh, O., Okunnu, G., Olarnpungoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Pioppner, F., Poindestre, A., Popovic, D., Primus, E., Pu, L., L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaeana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, X., Zhou, S., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 288118)
Worley, K.C.
Direct Submission
Submitted (27-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 288118)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRGM
Center clone name: CH230-293111
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 214989 bases at least Q40
Consensus quality: 222817 bases at least Q30
Consensus quality: 227320 bases at least Q20
Estimated insert size: 252471; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 5778: contig of 5778 bp in length
* 5779 5878: gap of unknown length
* 5879 16373: contig of 10495 bp in length
* 16374 16473: gap of unknown length
* 16474 90420: contig of 73947 bp in length
* 90421 90520: gap of unknown length
* 90521 236783: contig of 146262 bp in length
* 236783 236882: gap of unknown length
* 236882 250122: contig of 13240 bp in length
* 250122 250223: gap of unknown length
* 250223 257033: contig of 6811 bp in length
* 257033 257133: gap of unknown length
* 257133 259071: contig of 1938 bp in length
* 259071 259172: gap of unknown length
* 259172 261008: contig of 1837 bp in length
* 261008 261109: gap of unknown length
* 261109 264448: contig of 3340 bp in length
* 264448 264449: gap of unknown length

* 264549 267548: contig of 3000 bp in length
* 267548 267648: gap of unknown length
* 267648 271482: contig of 3834 bp in length
* 271482 271582: gap of unknown length
* 271582 275622: contig of 4040 bp in length
* 275622 275723: gap of unknown length
* 275723 286107: contig of 10385 bp in length
* 286107 286207: gap of unknown length
* 286207 288118: contig of 1911 bp in length.
* 288118 Location/Qualifiers
source
1. .288118
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-293111"
1. .1287
/notes="wgs end extension
clone end:T7"
1338. .5778
/notes="wgs end extension
clone end:T7"
3086. .3988
/notes="clone boundary
clone end:T7"
site:Mbol
end sequence:RXAIQ54TJB"
5879. .7959
/notes="wgs contig"
12280. .13622
/notes="wgs contig"
14019. .16373
/notes="wgs contig"
16474. .19067
/notes="wgs contig"
19118. .21281
/notes="wgs contig"
21332. .22964
/notes="wgs contig"
28345. .31644
/notes="wgs contig"
31695. .34111
/notes="wgs contig"
40505. .44913
/notes="wgs contig"
44964. .47869
/notes="wgs contig"
47920. .49820
/notes="wgs contig"
49871. .52672
/notes="wgs contig"
52723. .56843
/notes="wgs contig"
61887. .63717
/notes="wgs contig"
72165. .76565
/notes="wgs contig"
78769. .86319
/notes="wgs contig"
86370. .90420
/notes="wgs contig"
90521. .92438
/notes="wgs contig"
92489. .94959
/notes="wgs contig"
96097. .99542
/notes="wgs contig"
236883. .239542
/notes="wgs contig"

Query Match 48.2%; Score 841; DB 2; Length 288118;
Best Local Similarity 79.6%; Pred. No. 1.4e-123;
Matches 1083; Conservative 0; Mismatches 265; Indels 13; Gaps 7;


```

Db 44485 GCTACATGTACCCAGAACTGCATGTGCCAGAGACAGGCGCTGTTTCTTCGGGAAGCTT 44544
Qy 1348 CAGAACAAAGACTAACAGACTACTGTTGACATATACCTCCCGATGAACCAAGATT 1407
Db 44545 GAGAACAGAGACTACGGGGTACTGCTGGACATATATCTCCAGTGAACCAATGTG 44604
Qy 1408 GTGGGACACCAAGGTCATCTGTACCTCTGTGATGGGATGGCCCAAGATTCAGTTTTCGAG 1467
Db 44605 GAAGGCCACCAAGTCTTTTGTACTCTGCGCCAGGGATGGTCAGAACCAAGTTTTCGAG 44664
Qy 1468 TACACGTCCTCCAGAAAGAAATACGTTATACACCCACAGCGCTGAGGGCTGATGCTGTGTG 1527
Db 44665 TGTACATCTCCAGCAAGAAATACGTTATACCA - TCGCCAGGCCCAAGGTCGTGATACGACAGT 44723
Qy 1528 GAACAGCAATGGATACCTTATCATGCTCTGCGAAGAAATCGCCCAAGAAATCGTCCAG 1587
Db 44724 GAGGAAGGGAAGGAT - CCTTATCCCGGCTCTCGGAAGAAATACGTTCCAGAGAAATCAA 44781
Qy 1588 AAGTTTCATCTTGCAGGAGGATGATCTTTATTTTCCAGCAAGTCCCAAGAAATGTGTCAG 1647
Db 44782 GAGTTTCATCTTGCAGGAGGACCGACCTTAGTCTCAAGCAGAGCAGGAAATGTGTCAG 44841
Qy 1648 GCTCGAGGAGGAGTGCAGTGACAGTTTCCTTCACTTTACGAGACTGCACCAACTCG 1707
Db 44842 GCCACAGAGGAGGAGTCAACTCTGCTTTCGCAACTTGCAGAGCTGTAGCAACTCA 44901
Qy 1708 G--ATCATCAAAATGGTTCTTCTCAAGGCGCATGTTAT 1744
Db 44902 GACACAGCCAGAGGTGGTCTCAAGGAGTGCATGTGCAT 44940

RESULT 13
AC134060/c
LOCUS AC134060 166547 bp DNA linear HTG 22-SEP-2002
DEFINITION Rattus norvegicus clone CH230-227M12, *** SEQUENCING IN PROGRESS
***, 33 unordered pieces.
ACCESSION AC134060
VERSION AC134060.1 GI:23267331
KEYWORDS HTG; HTGS_PHASE1. (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 166547)
Murny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Bigwal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleaveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,I., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebrgeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gumaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kovis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Louisegh,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A.,

```

```

Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G.,
Olarpunsagoon,A., Pal,S., Parke,K., Pasternak,S., Paul,H.,
Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A.,
Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
Sisson,I., Sitter,C.D., Smajd,D., Sneed,A., Sodergren,E.,
Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
Svatek,A., Tabak,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
Walton,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczek,R.,
Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GTSQ
Center clone name: CH230-227M12
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 14248 bases at least Q40
Consensus quality: 149083 bases at least Q30
Consensus quality: 153531 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1215: contig of 1215 bp in length
* 1315: gap of unknown length
* 1316: contig of 1512 bp in length
* 2828: gap of unknown length
* 2928: contig of 1237 bp in length
* 4165: gap of unknown length
* 4264: contig of 1563 bp in length
* 4265: gap of unknown length
* 5828: contig of 2730 bp in length
* 5928: gap of unknown length
* 8658: contig of 1208 bp in length
* 8758: gap of unknown length
* 10065: contig of 2717 bp in length
* 12882: gap of unknown length
* 12893: contig of 1702 bp in length
* 14585: gap of unknown length
* 14684: contig of 2414 bp in length
* 17098: gap of unknown length
* 17198: contig of 1279 bp in length
* 17199:

```

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

```
* 18478 18577: gap of unknown length
* 18578 21168: contig of 2591 bp in length
* 21169 21268: gap of unknown length
* 21269 23708: contig of 2440 bp in length
* 23709 23808: gap of unknown length
* 23809 26530: contig of 2722 bp in length
* 26531 26631: gap of unknown length
* 26631 28801: contig of 2171 bp in length
* 28801 28901: gap of unknown length
* 28901 33091: contig of 4190 bp in length
* 33091 33191: gap of unknown length
* 33191 36968: contig of 3777 bp in length
* 36968 37068: gap of unknown length
* 37068 39304: contig of 2236 bp in length
* 39304 39404: gap of unknown length
* 39404 43433: contig of 4029 bp in length
* 43433 43533: gap of unknown length
* 43533 46236: contig of 2703 bp in length
* 46236 46336: gap of unknown length
* 46336 50699: contig of 4363 bp in length
* 50699 50799: gap of unknown length
* 50799 53947: contig of 3148 bp in length
* 53947 54047: gap of unknown length
* 54047 58293: contig of 4246 bp in length
* 58293 58393: gap of unknown length
* 58393 63160: contig of 4767 bp in length
* 63160 68334: contig of 5074 bp in length
* 68334 71462: contig of 5028 bp in length
* 71462 73562: gap of unknown length
* 73562 79617: contig of 6055 bp in length
* 79617 79717: gap of unknown length
* 79717 86568: contig of 6851 bp in length
* 86568 86668: gap of unknown length
* 86668 95652: contig of 8984 bp in length
* 95652 95752: gap of unknown length
* 95752 105300: contig of 9548 bp in length
* 105300 112938: contig of 7538 bp in length
* 112938 122447: contig of 9409 bp in length
* 122447 122548: gap of unknown length
* 122548 137894: contig of 15347 bp in length
* 137894 137995: gap of unknown length
* 137995 166547: contig of 28553 bp in length.

FEATURES             Location/Qualifiers
     source            1..166547
     /organism="Rattus norvegicus"
     /mol_type="genomic DNA"
     /db_xref="taxon:10116"
     /clone="CH230-227M12"

BASE COUNT  45529 a 35358 c 35370 g 45112 t 5178 others

Query Match      46.5%; Score 812.6; DB 2; Length 166547;
Best Local Similarity 79.2%; Pred. No. 4.5e-119;
Matches 1077; Conservative 0; Mismatches 269; Indels 13; Gaps 9;

QY 388  TATGATTATGATTAATTTGGCCAGAGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 447
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84617 TATGATTATGATTAATTTGGCCAGAGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 84558

QY 448  TCAACTCTCTTCCGAGCAGTTTACAGTGTCTTTCAGACATCCCGGATATCCTGCTAGAA 507
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84557 TCCACATCTCTTACAGCAGTTTACAGTGTCTTTCAGACATCTCCCGTATATCCTGCTGGAG 84498

QY 508  GAAGTGTATCTTTGATGATGACTACAGTGTATAGAGACACCTGAAGAGCGCTTGGCCAAAT 567
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84497 GAGTGTATCTTGTGTAGTACTACAGTGTACAGATCACCTGAAGAGCGCTTGGCCAAAT 84439

QY 568  GAGCTTTCCGGGACTGCGCAAGGTGCGCTGATCCGCGCCCAACAAGAGAGAGGCGCTGGTG 627
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
DB 84438 GAGCTGTACAGCTGCCCTAGGTGTACCTGATCCCTGGTAGCA--AGAGAGGGGCTTAGTG 84381
QY 628  CGAGCCCGGCTGTCTGGGGGGCTCTCGCGCGAGGGCGCATGTTCTGACATCTTCTCGACTGT 687
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84380 CAAG-CCGGCTACTGGGAGCCTCTGGGGCCAGGGGCGAAGTGTGCTGAGGTTTCTGCACTGT 84322
QY 688  CACTGTGAGTGCACAGGAAGGGTGGCTGGAGCCGCTGCTGCAGAGGATCCATGAAGAGAG 747
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84321 CACTGTGACTGTCTATGAGGGGTGGCTGGAG--TCTGCTGCAGAGGATCCAGAGAGAG 84264
QY 748  TCGGCACTGTGTGCCCGGTGATTGATGTGATGCTGACACTGGAACACCTTTCGATATCCTGGGG 807
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84263 TCGGCGGTGTGTGCCCGGTATCGATGTCACTTGTGAAACACTTCTGAGTACTCTGGGC 84204
QY 808  AACTCCGGGAGCCCCAGATCGCGGTTTCGACTGGAGGCTGGTGTTCACGTGCGACACA 867
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84203 AACTCC--GGAAGCCCCAGATTGGAGGCTTTGACTGGCGCTAGTATTACGTGGCAGCTG 84145
QY 868  GTTCTGAGAGGAGAGGATACGGATGCAATCCCCCGCTCGATGTCATCAGGTCTCCAAACA 927
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84144 GTCCCTTGGCGGAGCGGAAGTTGATGCGGACTCTATTGACATTATCAGGTCTCCAACT 84085
QY 928  ATGCTGTGCTGGCTGTTTGTGCTGTGAGTAAGAAATATTTTGAATATCTGGGGTCTTATGAT 987
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84084 ATGCTGGGGACTGTTTGTGTGAGTAAAGAGATATTTTGAATACCTGGGGTCTTATGAT 84025
QY 988  ACAGGAATGGAAGTTTGGGAGGAGAAAACTCGAAATTTTCTTTAGGATCTGGCAGTGT 1047
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84024 ACCGGAATGGAAGTCTAGGAGGAGAAAACTTGAATTTCTTTTAGGATCTGGCAATGT 83965
QY 1048 GGTGGGGTTCGTGAACACACACCATGTTCCCATGTTGGCCATGTTTTCCTCCCAAGCAAGCT 1107
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83964 GGTGACTCTCTGGAACACACACCTTACTCCAGCTGGGGCCACGCTTCTCCCTAAGCAAGCT 83905
QY 1108 CCTACTCCCGCAACAGGCTCTGGCCCAACAGTGTGCTGCAGCTGAAGTATGATGAT 1167
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83904 CCCTATTCACGAGCAAGGCTCTAGCCCAACAGTGTCCGAGCTGCAGAAAGTGTGGCGGAT 83845
QY 1168 GAATTTAAAGAGCTCTACTACCATCGCAACCCCGCTCCCGCTTGGAAACCTTTTGGGGAT 1227
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83844 GAATTTAAAGAGCTCTACCAACCCCGGAAATCCCC--AGCCCGCTTGGAAACCTCTGGGGAG 83786
QY 1228 GTGACAGAGAGGAAGCAGCTCCGGGCAACAGCTCCAGGTGTAAAGACTTCAAGTGTGTTCTTG 1287
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83785 CTGACAGAGAGAAAGAAAGCTTCAAGCCCAAGCTCCAGCTGTGAAGATTTAAAGCGGTTCTTG 83726
QY 1288 GAGCTGTGTATCCAGAACTGCGATGCTGCTGAGAGCAGGCTGGCTTCTTCGGGATGCTC 1347
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83725 GCTACACTGTACCCAGAACTGCAATGTGCCAGAGGACAGGCTGGTTTCTTCGGGAAAGCTT 83666
QY 1348 CAGAAACAAAGGACTTAAACAGACTACTGCTTTGACTATAAACCTCCCGATGAAACACAGATT 1407
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83665 GAGAACAGAGGACTACGGGGGTACTGCTGGACTATATATCTCCAGTGAAACAAATGTG 83606
QY 1408 GTGGGACACAGGTCATTCTGTACCTCTGTCTATGGGATGGGCCAGAAATCAGTTTTCGAG 1467
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83605 GAAGGCCACAGGCTCTTTTGTACCTCTGCGCCAGGATGGTTCAGAACCAAGTTTTCGAG 83546
QY 1468 TACAGCTCCAGAAAGAAATACGCTATAACACCCCAACAGCTGAGGGCTGATGCTGTG 1527
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83545 TGTACATCCAGCAAGAAATACGCTATAACA--TCGCCAGCCCGCAGGCTCTGCATAGCAGTG 83487
QY 1528 GAAGCAGGAATGATACCTTTATCATGCACTCTCGCAAGAAACTGCCCGCAGAGAAATCAG 1587
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83486 GAGAAAGGGAAGAT--CCTTATCCCGGGTCTCTGGAAGATATCCGTTCCAGAGAAATCAA 83429
QY 1588 AAGTTCATCTTGCAGGAGGATGGATCTTTATTTTTCAGAAACAGTCCAAAGAAATGTGTCAG 1647
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83428 GAGTTCATCTACAGGAGGACCGCACTTAGTCTAACAGCAGAGCAGGAATGTGTGGAG 83369
QY 1648 CTTGCGGAGGAAGAGTGTGAGTACAGTGTCTGTTTCCACTCTTTACGAGACTGCAACAACTCG 1707
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83368 GCCACAGAGGAGGAGTTCAACTCTGCTTTGCTCCGAACTTGGGAGACTGTAGCAACTCA 83309
```

JOURNAL

COMMENT

1708 G--ATCATCAGAAATGGTCTTCAAGAGCGCATGTTAT 1744
 |||||
 83308 GACACAGCCAGAGGTGGTCTTCAAGAGTGCATGTCAT 83270
 |||||

AC134060 166547 bp DNA linear HTG 22-SEP-2002
 Rattus norvegicus clone CH230-227M12, *** SEQUENCING IN PROGRESS
 ***, 33 unordered pieces.

AC134060
 AC134060.1 GI:23267331
 HTG; HTGS PHASE1
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

1 (bases 1 to 166547)
 Murzyn D.Marie, Metzker M.Lee, Abramson S., Adams C., Alder J.,
 Allen C., Allen H., Albrooks S., Amin A., Anguiano D.,
 Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,
 Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,
 Biswal K., Blair J., Blankenburg K., Blyth P., Brown M.,
 Bryant N., Buhay C., Burch P., Burrell K., Calderon B.,
 Cardenas V., Carter K., Cavazos I., Caesar H., Center A.,
 Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,
 Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza D.,
 Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,
 Delgado O., Denson S., Deramo C., Ding Y., Dinh H., Divya K.,
 Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,
 Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,
 Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,
 Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,
 Gbagregeorgis E., Geer K., Gill R., Grady M., Guerra W., Guevara W.,
 Gunaratne P., Haaland W., Hamil C., Hamilton C., Hamilton K.,
 Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,
 Hernandez R., Hines S., Hladun S.L., Hodgson A., Hognes M.,
 Hollins B., Howells S., Hulyk S., Hume J., Idlebird D., Jackson A.,
 Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,
 Karpasch S., Kelly S., Kelly S., Khan S., Khan S., Kovar C.,
 Kowis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,
 Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,
 Lorenshewa L., Louised H., Lozano R.J., Lu X., Ma J.,
 Maneshwar M., Mahindaratne M., Mamoud M., Malloy K., Mangum A.,
 Mangum B., Mapua P., Martin K., Martin R., Martinez E.,
 Mawhney S., McLeod M., McNeill T., Meenen E., Milosavljevic A.,
 Miner G., Minja E., Montemayor J., Moore S., Morgan M., Morris K.,
 Morris S., Munidasa M., Murphy M., Nair L., Nankervis C., Neal D.,
 Newton N., Nguyen N., Norris S., Nwaekelemeh O., Okwono G.,
 Olarnpungsoon A., Pal S., Parks K., Pasternak S., Paul H.,
 Perez A., Perez L., Pfannkuch C., Plopper F., Poindexter A.,
 Popovic D., Primus E., Pu L.-L., Puazo M., Quiroz J., Rachlin E.,
 Reeves K., Regier M.A., Reigh R., Reilly B., Reilly M., Ren Y.,
 Reuter M., Richards S., Riggs F., Rives C., Rodkey T., Rojas A.,
 Rose M., Rose R., Ruiz S.J., Sanders W., Savary G., Scherer S.,
 Scott G., Shatsman S., Shen H., Shetty J., Shvartsbeyn A.,
 Sison I., Sitter C.D., Smajs D., Sneed A., Sodergren E.,
 Song X.-Z., Sorelle R., Sosa J., Steimle M., Strong R., Sutton A.,
 Svatek A., Tabor P., Taylor C., Taylor T., Thomas N., Thomas S.,
 Tingey A., Trejos Z., Usmani K., Valas R., Vera V., Villasana D.,
 Waldron L., Walker B., Wang J., Wang Q., Wang S., Warren J.,
 Warren R., Wei X., White F., Williams G., Willson R., Wlezyk R.,
 Wooden H., Worley K., Wright D., Wright R., Wu J., Yakub S.,
 Yen J., Yoon L., Yoon F., Yu F., Zhang J., Zhou J., Zhou X.,
 Zhao S., Dunn D., von Niederhausern A., Weiss R., Smith D.R.,
 Holt R.A., Smith H.O., Weinstein G. and Gibbs R.A.

Direct Submission
 Unpublished
 Reference 2 (bases 1 to 166547)
 Rat Genome Sequencing Consortium.
 Direct Submission

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE

Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GTSO
 Center clone name: CH230-227M12
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 142428 bases at least Q40
 Consensus quality: 149083 bases at least Q30
 Consensus quality: 153531 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 33 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 1215: contig of 1215 bp in length
 1315: gap of unknown length
 1316: contig of 1512 bp in length
 1316: contig of 1512 bp in length
 2827: gap of unknown length
 2828: contig of 1237 bp in length
 4164: contig of 1237 bp in length
 4264: gap of unknown length
 4265: contig of 1563 bp in length
 5827: gap of unknown length
 5828: contig of 2730 bp in length
 5928: gap of unknown length
 6557: contig of 1208 bp in length
 6558: contig of 1208 bp in length
 8758: gap of unknown length
 8758: contig of 1208 bp in length
 9966: gap of unknown length
 10066: contig of 2717 bp in length
 12783: gap of unknown length
 12882: contig of 1702 bp in length
 12883: gap of unknown length
 14584: contig of 2414 bp in length
 14585: gap of unknown length
 14684: contig of 2414 bp in length
 14685: gap of unknown length
 17098: contig of 1279 bp in length
 17198: gap of unknown length
 18477: contig of 2591 bp in length
 18577: gap of unknown length
 21168: contig of 2591 bp in length
 21268: gap of unknown length
 21269: contig of 2440 bp in length
 21269: contig of 2440 bp in length
 23709: gap of unknown length
 23709: contig of 2722 bp in length
 26530: gap of unknown length
 26531: contig of 2171 bp in length
 28801: gap of unknown length
 28801: contig of 4190 bp in length
 28802: gap of unknown length
 28902: contig of 3777 bp in length
 33092: gap of unknown length
 33192: contig of 2703 bp in length
 36968: gap of unknown length
 36969: contig of 2236 bp in length
 39304: gap of unknown length
 39305: contig of 4029 bp in length
 39405: gap of unknown length
 43434: contig of 2703 bp in length
 43434: gap of unknown length
 46237: contig of 4363 bp in length
 46237: gap of unknown length
 50699: contig of 3148 bp in length
 50799: gap of unknown length
 53947: contig of 3148 bp in length
 53948: gap of unknown length
 54047: contig of 4246 bp in length
 58293: gap of unknown length
 58293: contig of 4246 bp in length
 58294: gap of unknown length

* 58394 63160: contig of 4767 bp in length
* 63161 63260: gap of unknown length
* 68334: contig of 5074 bp in length
* 68435 68435: gap of unknown length
* 73463 73462: contig of 5028 bp in length
* 73563 73562: gap of unknown length
* 79618 79617: contig of 6055 bp in length
* 79718 79717: gap of unknown length
* 86569 86568: contig of 6851 bp in length
* 86669 86668: gap of unknown length
* 95653 95652: contig of 8984 bp in length
* 95753 95752: gap of unknown length
* 105301 105300: contig of 9548 bp in length
* 105401 105400: gap of unknown length
* 112939 112938: contig of 7538 bp in length
* 113039 113038: gap of unknown length
* 122448 122447: contig of 9409 bp in length
* 122548 122547: gap of unknown length
* 137895 137894: contig of 15347 bp in length
* 137995 137994: gap of unknown length
* 166547: contig of 28553 bp in length.

FEATURES

source

1. 166547
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clones="CH230-227M12"

BASE COUNT 45529 a 35358 c 35370 g 45112 t 5178 others
ORIGIN

Query Match 46.1%; Score 805.4; DB 2; Length 166547;

Best Local Similarity 78.8%; Pred. No. 6.1e-118;
Matches 1073; Conservative 0; Mismatches 273; Indels 15; Gaps 9;

Qy 388 TATGATTATGATAATTTGGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 447
Db TATGATTATGATAATTTGGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGG 99957
Qy 448 TCAACTCTCTCCGACAGTTTACAGTGTCTCTGAGACATCCCGGATATCTCTAGAA 507
Db TCCACACTCTCTCAGACAG-TTACAGTGTCTCTGAGACTTCCCTGTATCTCTCTGGAG 100016
Qy 508 GAAGTGATCTTGTAGATGACTACAGTGATAGAGACACCTGAAGAGCGCTTGGCCAA 567
Db 100017 GAGTGTATTTGGGTAGATGACTACAGTGACAGAGATCACCTGAAGAGCGCTTGGCCAA- 100075
Qy 568 GAGCTTTCCGAGCTGCCAAGTCCGCTGATCCGCGCAACAGAGAGAGGGCTCTGTG 627
Db 100076 GAGCTGTACAGCTGCCNTAGGTGCACTGATCTCTCTAGCA--AGAGAGGGCTTAGT 100133
Qy 628 CGAGCCCGCTGTCTGGGGCGTCTCGCGGAGGGCGATGTTCTGACCTTCTCTGGACTGT 687
Db 100134 CAAG-CCGGCTACTGGGAGCCTTTGCGGCCAGGGCGAAGTGTCTGAGTTTCTGACTGT 100192
Qy 688 CACTGTGAGTCCACGAAGGTGTGCTGAGCCGCTGCTGCGAGAGGATCCATGAAGAGGAG 747
Db 100193 CACTGTGACTGTCTATGAGGGTGTCTGGAGTCCCTGCTGCGAGAGGATCCACGAAGAGG 100252
Qy 748 TCGGCACTGTGTCCCGGTGATTGATGTGATGCTGACCTGGAACACCTTCGAATACCTCGGG 807
Db 100253 TCGGCGGTGTGTGCCCGGTATCGATGTCTGACTGGAACACCTCTGAGTACTCTGGGC 100312
Qy 808 AACTCCGGGAGCCCCAGATCGCGGTGTTGACCTGGAGGCTGGTGTTCACGTGGCACACA 867
Db 100313 AACTCCAGAGAGCCCCAGATTGGTGGCTTTGACTGGCGCTAGTATTACAGTGGCAGTG 100372
Qy 868 GTTCTGAGAGGGAGAGATACGGAATGCCCGCTGATGTCATGTCATCAGGTCTCCAA 927
Db 100373 GTCCCTTGGGGAGCGGAAGTTGATCGGACTCCTATTGACATTTATCAGGTCTCCAACT 100432
Qy 928 ATGCTGCTGGCTGTTGCTGTAGTAAGAAATATTTTGAATATCTGGGGTCTTATGAT 987
Db 100433 ATGCTGCTGGGACTGTTGTTGTGTAGTAAGAGATATTTTGAATACCTGGGGTCTTATGAT 100492

Qy 988 ACAGGAATGAAGTTTGGGAGGAGAAAACTCGAAATTTTCTTTAGGATCTGCGACTGT 1047
Db 100493 ACCGGAATGAAGTCTGGGAGGAGAAAACTTGAATTTCTTTAGGATCTGCGACTGT 100552
Qy 1048 GGTGGGGTCTGGAAACACACACCCTATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCT 1107
Db 100553 GGT--GACTCAGGAACACACACCCTTACTCCACGTTGGGCCATTTCTTCCCTAAGCAAGCT 100610
Qy 1108 CCTTACTCCGCAACAAGGCTCTGGCCAAACAGTGTTCGTGAGCTGAAGTATGATGGAT 1167
Db 100611 CCCTATTACGCAAGGCTCTAGCCAAACAGTGTCCGAGTGCAGAAGTGTGACCGAT 100670
Qy 1168 GAATTTAAGAGCTCTACTACCATCGCAACCCCGTCCCGCTTGGAAACCTTTTGGGGAT 1227
Db 100671 GAATTTAAGAACTCTACCAACCCCGGAAACCCCG-AGCCCGCTTGGAAACCTCTTGGGGAG 100729
Qy 1228 GTGACAGAGAGGAGAGCTCCCGGCAACAGCTCCAGTGTAAGACTTCAAGTGGTCTTGT 1287
Db 100730 CTGACAGAGAGAAAGAGCTTCAGCCCAAGCTCCAGTGTGAAGACTTTAAAGCGGTTCTGT 100789
Qy 1288 GAGACTGTGTATCCAGAACTGTCATGTGCTGAGACAGGCTTGTCTTCCGGATGCTC 1347
Db 100790 GCTACNATGTACCAGAACTGTCATGTGCCAGAGACAGGCTTGTCTTCCGGAACTT 100849
Qy 1348 CAGAACAAAGAGCTTAAACAGACTACTGCTTTGACTATTAACCTCCCGATGAAGAACAGATT 1407
Db 100850 GAGAACACAGAGACGACGGGGGTACTGCTGGAAGTCTATAATCTCCAGTGAACACAGTGTGG 100909
Qy 1408 GTGGACACCAGGCTCATTTCTGTACCTCTGTGTCATGGGATGGCCAGATCAAGTTTTCGAG 1467
Db 100910 AAGGGCAACCCAGTGTCTTTGTATCTTGTGTCATGGGATGGTTCAGAACCAAGTTTTCGAG 100968
Qy 1468 TACACGTCCTCCAGAAAGAAATACGCTATAACACCCAGGCTGAGGGCTGCTGCTGTGTG 1527
Db 100969 TGTACATCTCCGCAAGAAATACGCTATTAACACCCAGGCTGAGGGCTGCTGCTGTGTG 101028
Qy 1528 GAAGCAGGAATGATACCTTTATCATGCTATCTCTGCAAGAAACTGCCCCAGAGAAATCAG 1587
Db 101029 GAGGAGGGAAGGAT--CCTTATTCGGATCTCTGGAAGATACGGTTCAGAGAAATCAA 101086
Qy 1588 AAGTTCATCTTGCAGGAGGATGA----TCTTTATTTACAGAACAGTCCAGAAATGTGT 1643
Db 101087 GAGTTCATCTTACAGGAGGACGACAGCTTAGTCTTACAGCAGACCGGGAATGTGT 101146
Qy 1644 CCAGGCTGCGAGGAAGAGTGCAGTGCAGTTCGTTCTTCCACTCTTACAGACTGCACAA 1703
Db 101147 GGAGGCCACAGAGAGAGTTCAACTCTGCTTGTCTCCAACTTTCGAGACTGTAGCAA 101206
Qy 1704 CTCGGATCATCAAAATGTTTCTTCAAGAGCGCATGTTAT 1744
Db 101207 CTCAGACAGCCAGAGGTGTTCTGCCAGGAGCTCATGTCTAT 101247

RESULT 15
AC098559 1/c
WPCOMMENT

Sequence split into 8 fragments LOCUS AC098559 Accession AC098559

Fragment Name	Begin	End
AC098559_0	1	110000
AC098559_1	100001	210000
AC098559_2	200001	310000
AC098559_3	300001	410000
AC098559_4	400001	510000
AC098559_5	500001	610000
AC098559_6	600001	710000
AC098559_7	700001	794616

Continuation (2 of 8) of AC098559 from base 100001 (AC098559 Rattus norvegicus clone CH24)

Query Match 46.0%; Score 803.8; DB 2; Length 110000;
Best Local Similarity 78.4%; Pred. No. 1.2e-117;
Matches 1065; Conservative 0; Mismatches 282; Indels 12; Gaps 8;


```
Qy 388 TATGATTATGATAAATTTGGCCAGGACATCTGTTATCATAGCATTTTAAATGAAGCCTGG 447
Db 55991 TATGATTATGATAACCTGGCCCAAGACATCCGTTGTTCATAGCATTTTAAATGAAGCCTGG 55932
Qy 448 TCAACTCTCCTTCGGACAGATTTACAGTGTCTTCAGACATCCCGGATATCCTGCTAGAA 507
Db 55931 TCCACACTCTCTCAGACAGATTTACAGTGTCTTCAGAGACTTCCCTGTATATCCTGCTGGAG 55872
Qy 508 GAAATGATCTCTGTAGATGACTACAGTGTATAGAGGACCTGAAGGAGCGCTTCGCCAAT 567
Db 55871 GAGGTCAATCTCTGTAGATGACTACAGTGTAGAGATCACCTGAAGGAGCGCTTCGCCAAG 55812
Qy 568 GAGCTTTCCGAGCTGCCCAAGGTGCGCTGATCGCGCAACAAGAGAGAGGCGCTTCGTTG 627
Db 55811 AGCTGTACACTGCCCTAGTGTACCTGTATCCTGTAGCA--AGAGAGGCGCTTAGTG 55754
Qy 628 CGAGCCCGCTGCTGGGGCGTCTGCGCGAGGGCGATGTTCTGACCTTCTCGACTGT 687
Db 55753 CAAG--CCGCTACTGGGAGCCTCTGGGCGCAGGGGCGAAGTGTCTGAGGTTTCTGACTGT 55695
Qy 688 CACTGTGAGTCCCAAGAGGCTGCGAGCGCTGTCGACAGAGATCCATGAAGAGGAG 747
Db 55694 CACTGTGACTGTGATGAGGGTGGCTGGAG--TCTGTGACAGAGATCCACGAGAGGAG 55637
Qy 748 TCGGCACTGTGTGCCCCGCTGATGATGATGACACTGGAACACCTTCGAATACCTGCGG 807
Db 55636 TCGGCGGTGTGTGCCCCGCTTATCGATGTATTGACTGGAACACTTCTGAGTACTTGGGC 55577
Qy 808 AACTCCGGGAGCCCCAGATCGGCGGTTTCACTGGAGGCTGGTTTCACTGCGACACA 867
Db 55576 AACTCC--GGAAGCCCCAGATTTGGAGGCTTGTACTGGCGCTTAGTATTACCTGGCAGGTG 55518
Qy 868 GTTCTGAGAGGAGAGATAGGATGCAATCCCGCTGATGTCATCAGGTCTCCCAACA 927
Db 55517 GTCCCTTGGCGGAGCGGAAGTTGATGCGGACTCTATTGACATTTACAGGTCTCCAACT 55458
Qy 928 ATGGCTGTGGCTGTTTGTGTGAGTAAGAAATATTTGAATATCTGGGGTCTTATGAT 987
Db 55457 ATGGCTGGGAGCTGTTTGTGTGATGATGATATTTGATATACCTGGGGTCTTATGAT 55398
Qy 988 ACAGAAATGGAATTTGGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGCGAGTGT 1047
Db 55397 ACCGGAATGGAATGCTAGGAGGAGAGAAACCTTGAATTTCTCTTTAGGATCTGCGAATGT 55338
Qy 1048 GGTGGGTTCTGGAAACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCT 1107
Db 55337 GGTGACTCTCTGGAACACACCCCTTACTCTCCACGTGGGCGACGCTCTTCCCTAAGCAAGCT 55278
Qy 1108 CCTACTCCCGCAACAAGGCTCTGGCCAAACAGTGTTCGTGAGCTGGAATGATGATGAT 1167
Db 55277 CCTATTTCAGGAGCAAGGCTCTAGCCAAACAGTGTCCGAGCTGAGAGTGTGGCGGAT 55218
Qy 1168 GAAATTAAGAGCTCTACTACATGCGAACCCCGCTGGCCCTTGGAACTTTTGGGGAT 1227
Db 55217 GAACTTAAGAACTCTACACCCCGAAATCCCC--AGCCCGCTGGAACCCCTCTGGGGAG 55159
Qy 1228 GTGACAGAGAGAGAGCAGCTCGGGAACAGCTCCAGTGTAAAGACTTCAAGTGGTCTTG 1287
Db 55158 CTGACAGAGAGAAAGAAAGCTTCAGCCCAAGCTCCAGTGTGAAGATTTAAAGCGGTTCCTG 55099
Qy 1288 GAGACTGTGTATCCAGAACTGCTGCTGAGGACAGGCTTCTTCGGGATGCTC 1347
Db 55098 GCTACACTGTACCCAGAACTGCTGATGTGCGAGGACAGGCTTGTCTTCGGGAGCTT 55039
Qy 1348 CAGAAACAGGACTTAAAGACTGCTGCTGTTGACTATTAACCTCCCGATGAAAAACAGATT 1407
Db 55038 GAGAAACAGGACTTACGGGGTACTGCTGCTGACTATTAATCTCCAGTGAACAAATGTG 54979
Qy 1408 GTGGACACACAGGTCAATCTGATCTGTCATGGGATGGGCGCAGATCACTTTTCGAG 1467
Db 54978 GAAAGCCACACAGGTCTTTGATCTCTGTCACCGGATGGGTCAGAACCACTTTTCGAG 54919
Qy 1468 TACACGTCCCAAGAAAGAAATAGCTTATAACACCCACAGCTGAGGGCTGCATTTGCTGTG 1527
```

Search completed: December 7, 2003, 18:37:52
Job time : 6377.45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 13:31:08 ; Search time 99.1527 Seconds
(without alignments)
7772.402 Million cell updates/sec

Title: US-10-074-527-3

Perfect score: 1746

Sequence: 1 atgtggggcgacggcg.....tcaaagagcgcatgttatga 1746

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	496.2	28.4	1737	4	US-09-217-306B-1
2	285.6	16.4	1920	3	US-09-376-856-1
3	284	16.3	3332	3	US-09-347-488-1
4	253.8	14.5	1521	4	US-09-795-926-30
5	253.8	14.5	1812	4	US-09-795-926-42
6	253.8	14.5	3896	4	US-09-795-926-44
7	232.2	13.3	1617	2	US-08-967-508-18
8	232.2	13.3	1617	3	US-08-967-506-18
9	232.2	13.3	1617	5	PT-US94-02552-18
10	232.2	13.3	1680	2	US-08-967-508-11
11	232.2	13.3	1680	3	US-08-967-506-11
12	232.2	13.3	1680	5	PT-US94-02552-11
13	232.2	13.3	2294	2	US-08-967-508-10
14	232.2	13.3	2294	3	US-08-967-506-10
15	232.2	13.3	2294	5	PT-US94-02552-10
16	222	12.7	3889	2	US-08-648-298-1
17	201.4	11.5	810	4	US-09-795-926-26
18	201.4	11.5	1101	4	US-09-795-926-38
19	198.8	11.4	1608	4	US-09-795-926-28
20	198.8	11.4	1899	4	US-09-795-926-40
21	134.8	7.7	675	4	US-09-795-926-22
22	134.8	7.7	966	4	US-09-795-926-34
23	97.6	5.6	903	1	US-08-620-312D-938
24	87.4	5.0	322	1	US-08-594-031-61
25	76.4	4.4	262	1	US-08-594-031-62
26	74.2	4.2	2416	4	US-08-620-312D-112
27	69.2	4.0	339	4	US-09-795-926-24

28	69.2	4.0	432	4	US-09-795-926-20	Sequence 20, Appl
29	69.2	4.0	630	4	US-09-795-926-36	Sequence 36, Appl
30	69.2	4.0	723	4	US-09-795-926-32	Sequence 32, Appl
31	64.6	3.7	331	3	US-09-328-111-94	Sequence 94, Appl
32	63.2	3.6	152331	3	US-09-128-155-16	Sequence 16, Appl
33	63	3.6	319	3	US-09-165-264-8	Sequence 8, Appl
34	60.6	3.5	320	3	US-09-165-264-11	Sequence 11, Appl
35	60.4	3.5	4488	3	US-08-406-030A-3	Sequence 3, Appl
36	59.8	3.4	320	3	US-09-165-264-14	Sequence 14, Appl
37	59.4	3.4	12001	1	US-08-458-568A-11	Sequence 11, Appl
38	59.2	3.4	2793	1	US-08-209-747-1	Sequence 1, Appl
39	59.2	3.4	2793	1	US-08-458-298-1	Sequence 1, Appl
40	59.2	3.4	4403765	3	US-09-103-840A-2	Sequence 2, Appl
41	59.2	3.4	4411529	3	US-09-103-840A-1	Sequence 1, Appl
42	58.8	3.4	2214	3	US-08-864-038A-1	Sequence 1, Appl
43	58.8	3.4	3331	3	US-08-864-038A-2	Sequence 2, Appl
44	58.8	3.4	3331	3	US-08-864-038A-4	Sequence 4, Appl
45	58.6	3.4	320	3	US-09-165-264-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1

US-09-217-306B-1
; Sequence 1, Application US/09217306B
; Patent No. 6465220
; GENERAL INFORMATION:
; APPLICANT: Hassan, Helle
; APPLICANT: Clausen, Henrik
; APPLICANT: Bennett, Eric P.
; TITLE OF INVENTION: Glycosylation Using Galnac-T4 Transferase
; FILE REFERENCE: 8850*1
; CURRENT APPLICATION NUMBER: US/09/217,306B
; CURRENT FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-217-306B-1

Query Match	28.4%	Score 496.2;	DB 4;	Length 1737;
Best Local Similarity	62.3%	Pred. No. 2.4e-104;		
Matches	796;	Conservative	0;	Mismatches 478;
Indels	3;	Gaps	1;	
Qy	220	GCCTGGCGCGCGGGCGAGCGGTGCGGCTGCAGCTGCAGGGCGAGGAGCTCGGGCTG	279	
Db	217	GCACCTTGGGAGTGGGGGAAGCCAGCAAACTCCAGCTCAACGAGGATGAATGAAGCAG	276	
Qy	280	CAGGAGGAGCGTGGCGCTGCACCAAGATTAACTACCTCAGCGACCCGCTCACTG	339	
Db	277	CAAGAAGAACTCAATTGAGATAGCCATCAATATTACCTCAGTGAAGATTCCCTG	336	
Qy	340	CACCGCGCTGCGCGAGCGCTGGAACCCGCTGCAAGAGAGAAATATGATTATGAT	399	
Db	337	CATCGACATAGAGGATAAAGATGATAGTGTAGTCCAGAGATCAACTATAGG	396	
Qy	400	AATTTGCCAGGACATCTGTTATCATAGCATTTTATTAATGAAGCCTGGTCAACTCTCTT	459	
Db	397	ACACTTCTACCACTCTGTTATCATGCTTTCTATAACGAAGCGTGGTCACTTCTGCTC	456	
Qy	460	CGGACAGTTTACAGTCTCTTGGACATCCCGGATATCTGCTAGAGAGATGATCTTT	519	
Db	457	CGTACCATTACAGTGTTTTAGAAACTTCTCTCGAGTCTTTTGAAGAGATCATCTTG	516	
Qy	520	GTAGATGACTACAGTATAGAGACACCTGAAGAGCGCTTGGCAATGAGCTTTCGGGA	579	
Db	517	GTGGATGACTTGGTGGACAGTTTATTGGAACACACAACTTGAACCTTACATCAGCAAT	576	
Qy	580	CTGCCCAAGTGGCGCTGATCCCGGCAACAAAGAGAGAGCGCTGTGCGAGCCGCGCTG	639	

577	CTTGATAGATGACGCTTGATTAGAGCAATAAGCAGAGGGCTGTTAGGGCCCGTCTG	636
640	CTGGGGCGCTCTGGCGAGGGCGATGTTCTGACCTTCTCGGACTCTCACTGTGAGTGC	699
637	ATTGGGGCCACTTTCGCCACTGGGACGTCCTCACTTCTCTGATTCTCACTGTGAGTGT	696
700	CACGAAGGTGGCTGGAGCGCGTCTCGCAGAGGATCCATGAAGAGGAGTGGCAGTGGTG	759
697	AAATCCGGTTGGCTGGAAACCGCTTTTGGAAAGGATTGGGAGATATGAAACAGCAGATTGTG	756
760	TGCCCGGTGATTGATGTCATCGACTGCGAAACACCTTTCGAATACCTGGGGGAACCTCGGGGAG	819
757	TGTCCTGTTATAGACACAATATGATTGGAATACCTTTGAAATCTATATGCGAGATAGGGGAG	816
820	CCCAGATCGCGCGTTCGACTGGAGGCTGGTGTTCACGTGGGCACACAGATTCCTGAGAGG	879
817	CCCATGATTGGTGGGTTTGACTGGCGTTTAACATTTCACTAGTGGCATTTCTGCCCAACACAG	876
880	GAGAGGATACGGATGCAATCCCCCGTCCGATGTCATCAGGTCCTCAACAATGGCTGTGGGG	939
877	GAAAGGACAGCGCGGATATCAAGAAATTTGACCCCATCAGATCACCTACATGGCTGGAGGA	936
940	CTGTTTGGCTGTGAGTAAGAAATATTTTGAATATCTGGGGTCTTATGATACAGGAATCGAA	999
937	CTGTTTGGCTGTGACCAAGAAATATTTTTCAGTACCTTGGAAACGTATGACACAGGAATCGAA	996
1000	GTTTGGGGAGGAGAAACCTCGAATTTTTCCTTTAGGATCTGGCAGTGTGGTGGGTTCTG	1059
997	GTTGTTGGGAGGTGAACACTTTGAGCTGTCTTTTAGGTGTGGCAGTGTGTGGCAAAATTG	1056
1060	GAAACACACCCCATGTTCCCATGTTGGCCATGTTTTCCCAAGCAAGCTCCCTACTCCCGC	1119
1057	GAGATCCACCCGTTTCCCAGTGGGCCATGTGTTCCCAAGCGGCACCATATGCTCGC	1116
1120	AACAAGGCTGTGGCCACACAGTGTTCGTGTGAGCTGAATATGATGGATGGATGAATTTAAAGAG	1179
1117	CCCAATTTCTTACAGAAATACTGCTCGGGCAGCAGAGTTTGGATGGATGAATACAAAGAG	1176
1180	CTCTACTACCATCCCAACCCCGTCCCGCTTGGAACTTTTGGGGATGTGACACAGAGG	1239
1177	CACTTCTACAATAGAAACCCCTCAGCAAGAAAGAGCTTATGTGTATTTTCTGAAAAGA	1236
1240	AAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGGTTCCTTGGAGAGCTGTAT	1299
1237	AAATTTACTACGAGCGGTTTGAGATGCAAGAGCTTTGACTGGTATTGAAAAACGTTTTT	1296
1300	CCAGAACTGCATGTGCTGTGAGGACAGGCCTGGCTCTTTCGGGGATGCTCCAGAACAAAGGA	1359
1297	CCTAATTTTACATGTTCCAGAGGATAGACACAGGCTGGCATGGGGCTATTTCGCAGTAGAGGG	1356
1360	CTAACAGACTACTGTTTGTACTATAACCTCCCGATGAAACACAGATTGTGGGACACAG	1419
1357	ATCTCGTCTGAATGTTTAGATTATAATTTCTCTGACAAACACCCACAGGTGTCTAAC--	1414
1420	GTCATTTCTGTAACCTGTGTCATGGGATGGGCCAGAAATCAGTTTTTTCGAGTACACGCTCCAG	1479
1415	-TTTCACTGTTTGGATGCCATGTGTCAGGAGGCAATCAATCTTTTGAATATACTTCAAC	1473
1480	AAAGAAATACGCTATAA	1496
1474	AAAGAAATAAGGTTTAA	1490

RESULT 2

```

US-03-376-856-1
; Sequence 1, Application US/093760856
; Patent No. 6235510
; GENERAL INFORMATION:
; APPLICANT: KUNWAR, SANJAY
; APPLICANT: VAN HORN, MARION M.
; APPLICANT: LARK, MICHAEL
; TITLE OF INVENTION: ppgAntase-T6
; FILE REFERENCE: GP-70638

```

```
Qy 1210 TTGGAACCTTTTGGGATGTGACAGAGGA-----AGCAGCTCCGGGACAAGCTC 1260
Db 1381 TCCTTGAGCAAGGCTGGAAGCCAGACTGCAATGGAAGCTTTGCAAGGAGACTG 1440
Qy 1261 CAGTGTAAAGACTTCAAGTGGTTCTTGAGAGCTGTGTATCCAGAACTGCATGTGCCTGAG 1320
Db 1441 GGTGTGGACATTCACATGGTTTCTGGCTAATGTCTACCTGAGCTGTACCCATCTGAA 1500
Qy 1321 GACAGGCTGGCTTTCTTGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGAC 1380
Db 1501 CCCAGGCCAGTTCTCTGGAAGCTCCACAACACTGGACT-----TGGGCTCTGT 1551
Qy 1381 TATAACCTCCCGATGAAACAGATTTGTGGGACACAGGTCATTTCTGTACTCTGTCT 1440
Db 1552 GCAGACTGCCAGGAGAGGGGACATCTGGGCTGTCCCATGGTGTGGCTCTCTGCACT 1611
Qy 1441 GGGATGGGCCAGAAATCAGTTTTCAGTACACGTCCCAAGAAAGAAATACGCTATAACACC 1500
Db 1612 GACAGCGGACACACAGTACTGACACACACAGCAGGAGGAGATTCATCTTTGGCAGC 1671
Qy 1501 CACAGGCTGAGGCTGCATTTCTGTGGAAGCAGGAATGGAATACCTTTATCATGCTCTC 1560
Db 1672 CCACAGC---ACCTGTGCTTGTCTCAGGCAGGAGCAGGTGATTCTTCAGAACTGCAAG 1728
Qy 1561 TGGGAAGAACTGCCCAGAGAAATCAGAAAGTTCAATCTTTCAGAGGAGTGGATCTTTATT 1620
Db 1729 GAGGAAGGCTTGGCCATCCACACAGCAGCACTGGGACTTCCAGGAGAAATGGGATGATGTC 1788
Qy 1621 CACGAACAGTCCAGAATGTGTCAGGCTCCGAGGAGGAGTGCAGTGACAGTTT 1676
Db 1789 CACATTTCTTGGGAATGATGGAAGCTGTGTGCAAGAAACAATAAAGATT 1844
```

RESULT 3

```
US-09-347-488-1
; Sequence 1, Application US/09347488
; Patent No. 6239266
; GENERAL INFORMATION:
; APPLICANT: Munroe, David
; APPLICANT: Pribill, Ingrid
; TITLE OF INVENTION: ZAP-3 TUMOR ASSOCIATED GENES AND THEIR
; TITLE OF INVENTION: USES
; FILE REFERENCE: GEN-2PRV
; CURRENT APPLICATION NUMBER: US/09/347,488
; CURRENT FILING DATE: 1999-07-02
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3332
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (368)...(2284)
US-09-347-488-1
```

```
Query Match 16.3%; Score 284; DB 3; Length 3332;
Best Local Similarity 53.8%; Pred. No. 9.6e-56;
Matches 740; Conservative 0; Mismatches 600; Indels 36; Gaps 6;
```

```
Qy 319 CTCAGCAGCCGATCTCACTGCAACCGCGCTGCCGAGCGCTGGAACCCGCTGTGAAA 378
Db 854 CTCAGTCCCGCATCCCTCCAGAGGCTCTGCCGAGGCTGCCGAGGCTGCTGTTCCATG 910
Qy 379 GAGGAGAAATATGATGATAATTTGCCAGGACATCTGTTATCATAGCAATTTTAAAT 438
Db 911 CTGAGCAGCAGCCCTCAGGACAGCTGCCACAGCCAGCGCTCATCTCTGTTTCCATGAT 970
Qy 439 GAAGCCTGGTCAACTCTCTCTCGGACAGTTTACAGTGTCTTTGAGACATCCCGGATATC 498
Db 971 GAGGCTTGGTCCATCTCTCTCGGACTGTACACAGCATCTTCGACACAGTGCACGGGCC 1030
```

```
Qy 499 CTGCTAGAAGAGTGTATCCTTTGTAGATGACTACAGTGTATAGAGAGCACTGAGAGGGC 558
Db 1031 TTCTGTAAGAGATCATCTCTGGACGACCTCAGCCAGCAAGGACAACCTCAAGTCTGCT 1090
Qy 559 TTGGCCAAATCAGCTTTTCGGGACTGCCCAAGGTGGCTGATCGCGCGCAACAAGAGAGAG 618
Db 1091 CTAGCGAATATGTGGCCAGGCTGGAGGGGTGAAGTTACTCAGGAGCAACAAGAGGCTG 1150
Qy 619 GGCTTGTGGAGCCCGGCTGCTGGGGCGCTCTCGCGCAGGGGCGATGTTCTGACCTTC 678
Db 1151 AGTGCCATCAGGCGCCGATGCTGGGGGCCACACAGAGCCACCGGGATGTCTCGTCTTC 1210
Qy 679 CTGCACTGTCTACTGTGATGCCAGGAAGGTGGCTGAGCGGCTGCTGAGAGATCAT 738
Db 1211 ATGGATCCCACTCGAGTGCACCCAGGCTGGCTGGAGCCCTCTCTCAGCAGAATAGCT 1270
Qy 739 GAAGAGGAGTCCGAGTGTGTCGGGTGATGATGATGATGATGATGATGATGATGATGAT 798
Db 1271 GGTGACAGGAGCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1330
Qy 799 TACCTGGGGAACCTCCGGGAGCCCCAGATCGGGGTTTTCGACTGGAGGCTGGTGTTCACG 858
Db 1331 TA---TTACCCCTCAAGGAGCTGCAGCGTGGGTGTTGGACTGGAAGCTGGATTTCCAC 1387
Qy 859 TGGCACAAGTCTCTGAGAGGAGAGATAGGATGCAATCCCGCTGCTGATGATCATCAGG 918
Db 1388 TGGGAACCTTTTGGCAGAGCATGTGAGGAAGGCCCTCCAGTCCCGCATAAAGCCCATCAGG 1447
Qy 919 TCTCCAAATAGTGGTGGGCTGTTGCTGTGATGATGATGATGATGATGATGATGATGATGAT 978
Db 1448 AGCCTGTGTGTCGGGAGAGGTGGTGGCCATGACACAGATTTACTTCCAAAAACACTGGA 1507
Qy 979 TCTTATGATACAGGAATGGAAGTTTGGGAGAGGAAACCTCGAATTTTCTTTTAGGATC 1038
Db 1508 GCGTATGACTCTCTTATGTCGCTCGAGGTGGTGAAGCTCGAATCTCTTCAAGGCC 1567
Qy 1039 TGGCAGTGTGGTGGGTTCTGGAACAACACCCATGTTCCCATGTTGGCCATGTTTTCCTCC 1098
Db 1568 TGGCTCTGTGGTGGCTCTGTTGAAATCCTTCCCTGCTCTCGGGTAGGACATCTACCAA 1627
Qy 1099 AAGCAGCTCCCTACTCCCGCAACAGGCTCTGGCC-----ACAGTGTTCGTGCA 1149
Db 1628 AATCAGGATTTCCATTTCCCTCCCTCGCAGGAGGCCACCTGAGGAACAGGTTTCGAT 1687
Qy 1150 GCTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1209
Db 1688 GCTGAGACCTGGCTGGGTCATTTCAAGAAACCTTCTACAGCATAGCCAGAGGCTTTC 1747
Qy 1210 TTGGAACCTTTTGGGATGTGA CAGAGAGGA-----AGCAGTCCGGGACAAGCTC 1260
Db 1748 TCCTTGAGCAAGGCTGAGAAGCCAGACTGATGGAACGCTTGACACTGCAAGAGAGACTG 1807
Qy 1261 CAGTGTAAAGACTTCAAGTGGTTCTTGGAGACTGTGTATCCAGAACTGCATGTCCTGAG 1320
Db 1808 GGTGTGCGGACATTCACATGGTTTCTGGCTAATGTCTACCTGAGCTGTACCCATCTGAA 1867
Qy 1321 GACAGCCTGGCTTCTTTCGGGATGCTCCAGAAACAAAGGACTAAACAGACTACTGCTTTGAC 1380
Db 1868 CCCAGGCCCAAGTTTCTCTGGAAGCTCCACAACACTGGACT-----TGGGCTCTGT 1918
Qy 1381 TATAACCTCCCGATGAAACACAGATTGTGGGACA CAGGTCATTTCTGTACCTCTGTCTAT 1440
Db 1919 GCAGACTGCCAGGAGAGGGGACATCTCGGCTGTCCCATGGTGTGGTCTCTTTCAGT 1978
Qy 1441 GGGATGGGCCAGAAATCAGTTTTCGAGTACACGTCCAGAAAGAAATACGCTATAACACC 1500
Db 1979 GACAGCGGAGCAACAGTACCTGCAACACACAGCAGGAAGGAGATTCTCTTTGGCAGC 2038
Qy 1501 CACAGGCTCAGGCTGCTATGCTGTGGAAGAGGAGGATGATACCTTTATCATGCTCTC 1560
Db 2039 CCACAGC---ACCTGTGCTTGTGTGTCAGGAGGAGGATGATTTCTTCAAGACTGCAAG 2095
Qy 1561 TGCAGAAAACTGCCCCAGAGAAATCAGAAAGTTTCTTTCAGGAGGATGATCTTTTATTT 1620
```

Db 2096 GAGGAAGCGCTGGCCATCCACAGCAGCTGGGACCTTCAGGAGAAATGGGATGTC 2155
Qy 1621 CACGAACAGTCCAGAATAATGTCTCCAGCTCGAGGAGAGTGCAGTGACAGTTT 1676
Db 2156 CACATTTCTTGGGAAATGCAATGCAAGCTGTGTGTCAGGAAGAAACAATAAAGATT 2211

RESULT 4

US-09-795-926-30
; Sequence 30, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 1521
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-795-926-30

Query Match 14.5%; Score 253.8; DB 4; Length 1521;
Best Local Similarity 55.5%; Pred. No. 5.9e-49;
Matches 561; Conservative 0; Mismatches 432; Indels 18; Gaps 3;
Qy 308 TTAACATCTACTCAGCGACCGCATCTCACTGCACCGCGCTGCGGAGCGCTGGAAAC 367
Db 47 TTAACATCTAGTCAGTGATTAATCTCTTGAATCGCTCTCCAGATATCGGCACC 106
Qy 368 CGCTGTGGAAGAGAAATATGATTAATGATTAATTTGCCAGGACATCTGTTATCATAG 427
Db 107 CAAATGCAACAGCAAGCGCTACCT--GGAGACATTTCCCAACACAGCATCATCATCC 163
Qy 428 CATTTTATAAGACCTGCTCAACTCTCTTCGGACAGTTTACAGTCTCTTACAGAT 487
Db 164 CTTTCCCAACAGAGGGCTGGTCTCTCTCCCGACCGTCCACAGTGTCTCAATCGCT 223
Qy 488 CCCCAGATATCTCTGCTAGAAGATGATCTCTTGTAGATGACTACAGTGTATAGAGACACC 547
Db 224 CGCTCCAGAGCTGCTCCCGAGATTGTTACTGTGCGACACTTTCAGTATCGAGACACC 283
Qy 548 TGAAGAGCGCTTGGCCAAATGAGCTTTCGGGACATGCCCCAAGTGGCCCTGATTCGCGCCA 607
Db 284 TGAAGAGCGCTTGAAGACTACATGGGCCCTTTTCCCGAGTGTGAGGATTTCTTGAACCA 343
Qy 608 ACAAGAGAGCGCTGTGTGGAGCGCGCTGCTGGGGCGCTCTCGCGGAGGGCGATG 667
Db 344 AGAAACGGGAAGGGCTGATAGGACCCGAAATGCTGGGGCGCTCACTGTGCAACTGGGATG 403
Qy 668 TTCTGACCTTCTGACTGTCACTGTGAGTGCACGAAGGGTGGCTGGAGCGCTGCTGTC 727

Db 404 TCATCACTTTTGGATTCACTGTGAAGCAATGTCAACTGGCTTCCCCCTTGGTTG 463
Qy 728 AGAGGATCCATGAAGAGGAGTGGCAGTGTGTGCCGGTGTGATGTGATCGACTGGA 787
Db 464 ACCGCAATGTCTCGGAACCGCAAGCAATGTGTGTGCCGATGATGATGATTAATGACCATG 523
Qy 788 ACACCTTTCAATACTCTGGGAACTCCGGGAGCGCCAGATCGGCGGTTTCGACTGGAGGC 847
Db 524 ACAGCTTTCGGTACGAGACACAGCAGGAGGATCCCATGCGGGAGCCTTTGACTGGGAGA 583
Qy 848 TGGTGTTCAGTGGCACACAGTTCTTCAGAGGAGAGGATACCGATGCAATCCCGCTGCG 907
Db 584 TGTACTACAAGCGGATCCCGATCCCT-----CCAGAACTGCAGAAAGCTGACCCACGCG 637
Qy 908 ATGTCAATCAGTCTCCCAACAAATGGCTGTGTGGGCTGTGTGCTGTGAGTAAGAAATATTTG 967
Db 638 ACCATTTTGAATCTCCCGTATGGCGGTGAGTGTGTGCGCGTGGATCGGAAGTGGTTCT 697
Qy 968 AATATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAGAAACCTCGAATTTT 1027
Db 698 GGGAACTCGCGGGTATGACCCAGGCTTGGAGATCTGGGAGGGGAGCAGTATGAATCT 757
Qy 1028 CCTTATAGGATCTGGCAGTGTGGGTTCTTGGAAACACACACCATGTTCCCATGTTGGCC 1087
Db 758 CCTTCAAGGTGTGGATGTGTGGGGCGCATGAGGAGCATCCCTGCTCCAGGGTGGGCC 817
Qy 1088 ATGTTTTTCCCAAGCAAGCTCCCTAC-----TCCGCAACAAAGGCTCTGCCCAACA 1138
Db 818 ATATCTACAGGAAGTATGTGCCCTCAAGGTCCCGCGCGAGTCAAGCTGGCCCGGAAACC 877
Qy 1139 GTGTTCTGTCAGTGAAGTATGATGATGATGAAATTTAAAGAGCTCTACTACCATCGCAACC 1198
Db 878 TTAAGCGGTGGCGGAAGTGTGGATGGATGAGTACGACAGATACATTTACAGCGCGCGC 937
Qy 1199 CCGTGTCCCGCTTGGAAACCTTTTGGGATGTGACAGAGAGAGCAGCTCGGGAACAAGC 1258
Db 938 CTGAATACCGCCACCTCTCCGTGGGATGTGCGAGTCCAGAAAAAGCTCCGAGCTCCC 997
Qy 1259 TCCAGTGTAAAGACTTCAAGTGGTTCTTGGAGACTGTGTATCCAGACTGC 1309
Db 998 TTAACCTGCAAGAGTTCAGTGGTTTATGACGAAGATAGCCTGGGACCTGC 1048

RESULT 5

US-09-795-926-42
; Sequence 42, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 42

; LENGTH: 1812

; TYPE: DNA

; ORGANISM: homo sapiens

US-09-795-926-42

Query Match 14.5%; Score 253.8; DB 4; Length 1812;

Best Local Similarity 55.5%; Pred. No. 6.3e-49;

Matches 561; Conservative 0; Mismatches 432; Indels 18; Gaps 3;

```
Qy 308 TTAACATCTACTCTCAGCGACCGCTCTCACTGACCGCGCTCCCGAGCGCTGGAAACC 367
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 338 TTAACATCTAGCTCAGTGATAAATCTCTTGAATCGCTCTCTCCAGATATCCGGACCC 397
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 368 CGCTGTGCAAGAGAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAG 427
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 398 CAAACTGCAACAGCAAGCGCTACCT--GGAGACACTTCCCAACACAGCATCATATCC 454
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 428 CATTTTATATGAAGCCTGCTCAACTCTCTTCGGACAGTTTACAGTGTCTTGAGACAT 487
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 455 CTTTCCAAACGAGGGGTGGTCTCTCTCTCCGACCGTCCACAGTGTGCTCAATCGCT 514
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 488 CCCCGATATCTCTAGAAAGTATCTCTGTAGATGACTACAGTATAGAGACACC 547
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 515 CGCTCCAGAGCTGTGCGCGAGATTGACTGTGTCGACACTTCAGTATCGAGAGACCC 574
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 548 TGAAGAGCGCTTGCCCAATGAGCTTTCGGGACTGCCCAAGGTGCGCTGATCCGCGCCA 607
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 575 TGAAGAGCCTTTGAAGACTACATGCGCTTTTCCCGAGTGTGAGGATCTTCGAACCA 634
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 608 ACAAGAGAGGGGCTGTGCGAGCCCGCTGTCTGGGGGCTGTGCGCGAGGGCGCATG 667
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 635 AGAAACGGGAGGGCTGATAAGGACCCGAATGCTGGGGGCTCAGTGGCAACTGGGATG 694
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 668 TTCTGACTTCTCGACTGTACACTGTAGTCCACGAGGCTGCTGAGCGGTGCTGTC 727
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 695 TCATCAATCTTGGATTACACTGTGAAGCAATGTCAACTGGCTTCCCGCTTGTGTTG 754
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 728 AGAGGATCCATGAAGAGGAGTGGCGAGTGTGTCGGGTGATTGATGTGATCGACTGGA 787
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 755 ACCGATTTGCTCGGAACCGCAAGACCATTTGTGTCCCGATGATTGATGAATGACCATG 814
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 788 ACACCTTTCGAATACCTGGGGAACCTCCGGGGAGCCCCAGATCGCGGTTTCGACTGGAGGC 847
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 815 ACAGCTTTCGGTACGAGACACAGGACGGGATGCCATGCGGGAGCCTTTGACTGGGAGA 874
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 848 TGGTGTTCACGTGGCACACAGTTCTGAGAGGAGAGATACGATGCAATCCCGCTGCG 907
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 875 TGTACTACAAGCGGATCCCGATCCCT-----CCAGAACTGCAGAAAGCTGACCCAGCG 928
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 908 ATGTTCATCAGGTCTCCAAACAATGCTGTGGGCTGTTTGTCTGAGTAAGAAATATTTTG 967
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 929 ACCATTTGATGCTCCCGTGTATGCGCGGTGACTGTTGCGCGTGGATCGGAATGGTTCT 988
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 968 AATATCTGGGTCTTATGATAAGAAATGGAAGTTTGGGGAGGAGAAACCTCGAATTTT 1027
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 989 GGGAACTCGGCGGATGACCCAGGCTTGGAGATCTGGGAGGGGAGCAGTATGAATCT 1048
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1028 CCTTTAGATCTGCGAGTGTGGGGTTCTGGAAACACACCCATGTTCCATGTTGGCC 1087
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1049 CCTTCAAGGTGTGATGTGTGGGGGCGCATGAGGACATCCCTCTGCTCCAGGGTGGGCC 1108
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1088 ATGTTTTTCCCAAGCAAGCTCCCTAC-----TCCCGCAACAAGCTCTGCGCAACA 1138
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1109 ATATCTACAGGAATGATGTGCCCTTACAGGTCCCGCGGGAGTCAAGCTGCGCGCAACC 1168
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1139 GTGTTCTGCGAGCTGAAGTATGGAATGGAATTTAAAGAGCTCTACTACCATCGCAACC 1198
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1169 TTAAGCGGGTGGCGAAGTGTGGATGATGATGACGAGAGTACATTTACCAGCGCGGC 1228
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1199 CCGTGCCTGTTGGAACTTTTGGGATGTGACAGAGAGGAACAGCTCTCGGGAACGC 1258
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 1229 CTGAATACCGCACCTCTCCGCTGGGATGTCGAGTCCAGAAAAAGCTCCGAGCTCCC 1288
Qy 1259 TCCAGTGTAAAGACTTTCAGTGGTTCTTGGAGACTGTGTATCCAGAACTGC 1309
Db 1289 TTAACGCAAGAGTTTCAAGTGGTTTATGACGAAGATAGCTGGGACCTGC 1339
```

RESULT 6

US-09-795-926-44

; Sequence 44, Application US/09795926

; Patent No. 6555669

; GENERAL INFORMATION:

; APPLICANT: Donoho, Gregory

; APPLICANT: Hilbun, Erin

; APPLICANT: Turner, C. Alexander Jr.

; APPLICANT: Friedrich, Glenn

; APPLICANT: Abuin, Alejandro

; APPLICANT: Zambrowicz, Brian

; APPLICANT: Sands, Arthur T.

; APPLICANT: Walke, D. Wade

; APPLICANT: Wilganowski, Nathaniel L.

; APPLICANT: Hu, Yi

; APPLICANT: Kieke, James Alvin

; APPLICANT: Potter, David George

; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

; FILE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME

; FILE REFERENCE: LEX-0144-USA

; CURRENT APPLICATION NUMBER: US/09/795,926

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: US 60/185,920

; PRIOR FILING DATE: 2000-02-29

; PRIOR APPLICATION NUMBER: US 60/186,558

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: US 60/191,849

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 47

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 44

; LENGTH: 3896

; TYPE: DNA

; ORGANISM: homo sapiens

US-09-795-926-44

Query Match

Best Local Similarity 14.5%; Score 253.8; DB 4; Length 3896;

Matches 561; Conservative 0; Mismatches 432; Indels 18; Gaps 3;

```
Qy 308 TTAACATCTACTCTCAGGACCGCATCTCACTGACCGCGCTGCGCGAGCGCTGGAAACC 367
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 347 TTAACATCTACGTACGTGATATAAATCTCTTGAATCGCTCTCTCCAGATATCCGGACCC 406
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 368 CGCTGTGCAAGAGAGAAATATGATTATGATATTTGCCAGGACATCTGTTATCATAG 427
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 407 CAACTGCAACACAGACGCGTACCT---GGAGACACTTCCCAACACAGATCATCATCC 463
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 428 CATTTTATATGAAGCCTGTGCTCAACTCTCTCTCGGACAGTTTACAGTGTCTTGAGACAT 487
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 464 CCTTCCACACAGAGGGCTGTCTCTCTCTCTCCGACCGTCCACAGTGTCTCAATCGCT 523
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 488 CCCCGATATCTGCTAGAAAGAGTGTATCTTGTAGATGACTACAGTGTAGAGAGACCC 547
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 524 CGCCTCCAGAGCTGGTCCCGAGATGTACTGGTTCGACGACTTCAGTGTAGAGAGACCC 583
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 548 TGAAGAGCGCTTGGCCCAATGAGCTTTCGGGACATGCGCAAGGTGCGCTGATCCGCGCCA 607
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 584 TGAAGAGCGCTTGTGAAGACTCATGGCCCTTTTCCCGAGTGTGAGGATTTCTTGAACCA 643
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 608 ACAAGAGAGAGCGCTGGTTCGAGCGCGCTGTCTGGGGGCGTCTGCGCGGAGGGCGGATG 667
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 644 AGAAACGGGAAGGGCTGATGAAGACCCGATGCTGGGGGCTCAGTGGCAACTGGGGATG 703
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 668 TTCTGACTTCTGGAATCTTGGGATGTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 727
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 704 TCATCATTCTTGATTCACACTGTGAAGCAATGTCACTGCTTCCCTTGGCTG 763
Qy 728 AGAGATCCATGAAGAGAGTGGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 787
Db 764 ACCGATTGCTCGGAACCGCAAGACCATGTGTGCGCGATGATGATGATGATG 823
Qy 788 ACACCTTCGATACCTCGGGAACCTCGGGAAGCCCGAGATCGGCGGTTTCGAC 847
Db 824 ACGATTTCGTACGACACAGCAGGAGGGATGCCATGCGGGAGCCCTTGAC 883
Qy 848 TGGTGTTCACGTGGCAGACAGTTCCTGAGAGGAGAGATACGATGCAATCCCG 907
Db 884 TGTACTACAGCGGATCCGATCCCT-----CCAGAACTCAGAAAGCTGACCC 937
Qy 908 ATGTCATCAGCTCCCAACATGCTGTGGGCTGTGCTGCTGAGTAAAGAAATTT 967
Db 938 ACCATTGAGTCTCCCGTGATGCGCGGTGACTGTTGCGCGTGATCGGAAGTGT 997
Qy 968 AATATCTGGGTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAGAAACCT 1027
Db 998 GGAACCTCGCGGATGACCCAGGCTTGGAGATCTGGGAGGGGAGCAGTATCA 1057
Qy 1028 CTTTAGATCTGGCAGTGGTGGTCTGGAACACACACCAATGTTCCATGTTGCC 1087
Db 1058 CTTCAAGTGTGGATGTGGGGCCCGCATGAGAGACATCCCTGCTCCAGGCTGG 1117
Qy 1088 ATGTTTCCCAAGCAGCTCCCTAC-----TCCGCAACAGGCTCTGCCAACA 1138
Db 1118 ATATCTACAGGAATGATGCGCCTACAGGTCCCGCGCGGAGTCAGCGTGGCC 1177
Qy 1139 GTGTTCTGTCAGCTGAAGTATGGATGGAATTTAAAGAGCTCTACTACCACG 1198
Db 1178 TTAAGCGGTGGCGAAGTGGATGGATGATGATGATGATGATGATGATGATGAT 1237
Qy 1199 CCGTGGCCGTGGAACTTTTGGGATGTGACAGAGAGAGAGAGAGAGAGAGAG 1258
Db 1238 CTGAATACCGCACCTCTCCGCTGGGATGTGCGAGTCCAGAAAGCTCCGAGCT 1297
Qy 1259 TCCAGTGAAGACTTCAAGTGGTCTTGGAGCTGTATCCAGACTGC 1309
Db 1298 TTAAGTGAAGGTTTCAAGTGGTATGACGAAGATAGCTGGGACCTGC 1348
```

RESULT 7

```
US-08-967-508-18
; Sequence 18, Application US/08967508
; Patent No. 5910570
; GENERAL INFORMATION:
; APPLICANT: The Upjohn Company
; APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GALNAC:
; TITLE OF INVENTION: Polypeptide, N-Acetylglucosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual
; ADDRESS: Property Legal Services
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,508
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/602,830
; FILING DATE: 13 No. 5910570ember 1995
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755. P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-833-2210
; TELEFAX: 616-833-8897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-967-508-18
```

```
Query Match 13.3%; Score 232.2; DB 2; Length 1617;
Best Local Similarity 54.8%; Pred. No. 5.2e-44;
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;
```

```
Qy 392 ATTATGATTAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAAGCCTGGTCAA 451
Db 269 ATCAGATAACCTTCCTACAAACAGTGTGTGTGTTTCCCAATGAGGCTTGGAGCA 328
Qy 452 CTCTCTCTTCGACAGTTTACAGTGTCTTCCAGACATCCCGGATATCCTGCTAGAAAG 511
Db 329 CACTTCTGCAATGTCATAGGTCATTAATCGCTCACCAGGCACATGCTAGAGAA 388
Qy 512 TGATCTCTTGTAGTACTACAGTATAGAGACACCTGAAAGGAGCGGTTGCCCAATGAGC 571
Db 389 TTGTTCTAGTAGATGATGCGCAGTGAAGAGACTTTTTTAAAGAGACTCTTAGAGAGTTACG 448
Qy 572 TTTGCGGACT---GCCCAAGTGGGCTGATCGCGCCACCAAGAGAGAGGCGCTGGTGC 628
Db 449 TGAATAAATTTAAAGTACCCGTTTCAAGTCAATCGAATGGAGAGCGTTCTGGATTGATCA 508
Qy 629 GAGCGCGGCTGCTGGGCGCTCTGCGCGAGGGCGGATGTTCTGACCTCTCTGAGCTGTC 688
Db 509 GAGCTAGTTAAAGTGTGCTGTCTAAAGCCCAAGTGTACCTTTTGTAGACCGCG 568
Qy 689 ACTGTGAGTCCCAAGAGGTTGGAGCGGCTGTCGAGAGAGATCCATGAAGAGAGT 748
Db 569 ACTGTGAGTGCACAGTGGGCTGGAGCGCTCTCTTAGCCAGGATCAAAACATGACAGGA 628
Qy 749 CGGAGTGGTGTGCGCGGTTTCGACTGGAGGCTGGTGTTCACGTTGGGACACAG 808
Db 629 AGACAGTGGTGTGCTCCCATCATAGATGTGATCAGTGAATGACACTTTTCGAGTACATGGCAG 688
Qy 809 ACTCCGGGAGCGCCAGATCGGCGGTTTCGACTGGAGGCTGGTGTTCACGTTGGGACACAG 868
Db 689 GTTCTGACATGACCTA---TGGCGGTTCACTGGAAGCTCACTTTTCGCTGGTATCCTG 745
Qy 869 TTCTGTAGAGGAGAGAGG---ATAAGGATGCAATCCCGGCTGATGTGTCATCAGGTCTCAA 925
Db 746 TTCCCAAGAGAAATGGACAGAAAGAGTGTGATCGGACTCTCTCTGTGTAGAAACACCTA 805
Qy 926 CAATGGCTGTGGCTGTTTGTGTGTGATGAAGAAATATTTTGAATATCTGGGGTCTTATG 985
Db 806 CAATGGCAGGAGGCGCTTTTTCATATAGACAGAGATTACTTTCAGGAAATTTGGAAACATATG 865
Qy 986 ATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAAATTTTCTTTAGGATCTGGGAGT 1045
Db 866 ATGCTGGAATGGATATTTTGGGAGGAGAAACCTAGAAATTTCTTTAGGATTTGGCAGT 925
Qy 1046 GTGTGGGCTTCTGGAAACACACCCATGTTCCCATGTTGGCCATGTTTTCGCCAAGCAAG 1105
Db 926 GTGGAGGAACCTTGGAGATGTTTACTTGTCTCACATGTTGGACATGTGTTCGGAAGAGCTA 985
Qy 1106 CTCCCTACTCCGCAACAGGCTCTGGCCACAGTGT-----TCGTGCGAG 1150
Db 986 CACCTTACAGGTTTCCAGGAGGCGCGGGCAGATTATCAATAAAAAATAACAGACGACTTG 1045
```


QY 1151 CTGAAGTATGATGGATCAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCT 1210
DB 1046 CAGAAGTATGATGGATGAATTTCAAGAAATTCCTCTATATATTAATTTCCAGGTGTACAA 1105
QY 1211 TGAACCTTTTGGGGATGTGACAGAGAGGAGCAGCTCCGGGACAAGCTCCAGGTGTAAG 1270
DB 1106 AGGTAGATTAGAGATATATCATCAAGACTTGGCTTAAGGCACAACTCCAATGCAGAC 1165
QY 1271 ACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCAATGTCCTGAGACAGGCGCTG 1330
DB 1166 CATTCCTCTGGTACTAGAGAAATTTATCTGATTCTCAGATTCCCTCGTCACTATTCT 1225
QY 1331 GCTTCTTCGGGATCTCCAGAACAAAGACTAAACAGACTACTGTTTGACTATACCCCTC 1390
DB 1226 CTTTGGGAGAGATAC-----GAAATGTGGAACAAATCAAGTGTCTAGATAACATGCTA 1279
QY 1391 CCGATGAAACCAAGATTGTGGGACACCCAGGTCAATTCGTACTCTGTGATGGGATGGCC 1450
DB 1280 GAAAGAGATGAAAGTTGGA-----ATTTTAACTGTCATGTTGGATGGAG 1327
QY 1451 AGAATCAGTTTTCGAGTACACGTCGCCAGAAAGAAAT 1487
DB 1328 GTAATCAGGTTTCTCTACACTGCCAACCAAGAAAT 1364

RESULT 8

US-08-967-506-18
; Sequence 18, Application US/08967506
; Patent No. 6096512
; GENERAL INFORMATION:
; APPLICANT: The Upjohn Company
; APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAC:
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual
; ADDRESSEE: Property Legal Services
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,506
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/602,830
; FILING DATE: 13 No. 6096512ember 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Barnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-833-2210
; TELEFAX: 616-833-8897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-967-506-18

Query Match 13.3%; Score 232.2; DB 3; Length 1617;

Best Local Similarity 54.8%; Pred. No. 5.2e-44;
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;
QY 392 ATTATGATATTTGCCAGGACATCTGTTATCATAGCATTTTATAAGACCTGGTCAA 451
DB 269 ATCCAGATACTTCTTACCAACAGTGTGTGATTTTCCCAATAGGCTTGGAGCA 328
QY 452 CTCTCCTTCGACAGTTTACAGTGTCTTGGAGATATCCCGGATATCTCTCTAGAAAG 511
DB 329 CACTTCTCGAACTGTCCATAGCGTCATTAATCGTCCACCAAGGCACATGCTAGAGAAA 388
QY 512 TGATCCTTGTAGATGACTACAGTATAGAGAGCACCTGAAGGAGCGCTTGGCCATAGC 571
DB 389 TTGTTCTAGTAGATGATGCCAGTGAAGAGACTTTTAAAAAAGACCTCTAGAGAGTTACG 448
QY 572 TTTCCGGACT---GCCAAGGTGCGCTGATCCGCGCAACAAGAGAGAGAGCGCTGTGC 628
DB 449 TGAATAAATTTAAAGTACCCGTTTCAGTCATTGAATGGAGCAGCGTCTGGATTGATCA 508
QY 629 GAGCCGCGTGTCTGGGGCGTCTGCGCGAGGGCGGATGTTCTGACCTTCTCGACTGTC 688
DB 509 GAGCTAGTTTAAAAAGGTGCTGTGTCTAAAGGCCAAGTATCACCTTTTACGCGC 568
QY 689 ACTGTGATGTCACAGTGGGCTGAGAGGCTGTCGAGAGGATCCATGAAGAGAGT 748
DB 569 ACTGTGATGTCACAGTGGGCTGAGAGGCTGTCGAGAGGATCCATGAAGAGAGT 628
QY 749 CGCAGTGGTGTCCCGGTCATTGATGATGATGATGATGATGATGATGATGATGATGATG 808
DB 629 AGACAGTGTCTGCTCCATCATAGATGATGATGATGATGATGATGATGATGATGATG 688
QY 809 ACTCCGGGGAGCCCGAGATCGGCGGTTTTCGATGGAGGCTGGTGTTCACGCTGGCACACAG 868
DB 689 GTTCTGACATGACCTA---TGGCGGGTTCACTTGAAGCTCAACTTTCGCTGGTATCTG 745
QY 869 TTCTGAGAGGAG 925
DB 746 TTCCCAAGAGAGAAATGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 805
QY 926 CAATGGCTGGTGGCTGTTTGTGCTGAGTGAAGAAATATTTTGAATATCTCGGGTCTTATG 985
DB 806 CAATGGCAGGAGGCTTTTCAATAGACAGAGATTACTTTCAGGAATTTGGAACATATG 865
QY 986 ATACAGGAATGGAAGTTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1045
DB 866 ATGCTGGAATGGATATTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 925
QY 1046 GTGCTGGGTTCTGGAAACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAG 1105
DB 926 GTGGAGGAACCTTTGGAGATGTTTACTTGTCTCATATGTTGACATGTTTTCGGAAGCTA 985
QY 1106 CTCCCTACTCCCGCAACAGGCTCTGGCCACAGTGT-----TCGTGACAG 1150
DB 986 CACCTTACAGCTTTCCAGGAGGACCGGGCAGATTTATCAATAAAATAACAGACAGCTTG 1045
QY 1151 CTGAAGTATGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTCCCGCT 1210
DB 1046 CAGAAGTATGATGGATGAATTTCAAGAAATTTCTCTATATAATTTCTCCAGGTGTACAA 1105
QY 1211 TGAACCTTTTGGGGATGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1270
DB 1106 AGGTAGATTATGGAGATATATCATCAAGACTTGGCTTAAGGCACAACTCCAATGCAGAC 1165
QY 1271 ACTTCAAGTGGTCTTGGAGACTGTCTATCCAGAACTGATGTCCTGAGGAGAGGCGCTG 1330
DB 1166 CATTCCTTGGTACCTAGAGAAATTTATCTGATTCTCAGATTCCCTCGTCACTATTCT 1225
QY 1331 GCTTCTTCGGGATGCTCCAGAACAAAGGACTTAACAGACTACTGCTTGTGACTATAACCCCTC 1390
DB 1226 CTTTGGGAGAGATAC-----GAAATGTGGAACAAATCAGTGTCTAGATAACATGCTA 1279
QY 1391 CCGATGAAACCAAGATTGTGGGACACCCAGGTCAATTCGTACTCTGTGATGGGATGGCC 1450

Db 1280 GAAAGAGAAATGAAAAAGTTGGA-----ATTTTAACTGTCTATGGATGGAG 1327
Qy 1451 AGAATCAGTTTTCAGTACACGTCGCCAGAAAGAAAT 1487
Db 1328 GTAATCAGGTTTCTCTTACACGTCGCCAAAGAAAT 1364

RESULT 9
PCT-US94-02552-18
; Sequence 18, Application PC/TUS9402552
; GENERAL INFORMATION:
; APPLICANT: Elhammer, Ake P.
; APPLICANT: Homa, Fred L.
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNac:
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Upjohn Company, Corp. Intellectual
; ADDRESSEE: Property Law
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/02552
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-385-5210
; TELEFAX: 616-385-6897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US94-02552-18

Query Match 13.3%; Score 232.2; DB 5; Length 1617;
Best Local Similarity 54.8%; Pred. No. 5.2e-44;
Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;

Qy 392 ATTATGATAATTTGCCAGGACATCTGTATCATAGCAATTTTATAATGAAGCCCTGGTCAA 451
Db 269 ATCCAGATTAACCTTCTCAACACAGTGGTGGTATTTTCCACATGAGGCTTGGAGCA 328

Qy 452 CTCCTCTTCGGACAGTTTACAGTGTCTTTGAGACATCCCGGATATCCTGCTAGAAGAAG 511
Db 329 CACTTCTCGAACTGTCTCATAGCGTCAATTAATCGCTCAACCAAGGCACATGCTAGAAGAA 388

Qy 512 TGATCCTTGTAGATGACTACAGTGTAGAGACACCTGAAGAGCGCTTGCCCATGAGC 571
Db 389 TTGTTCTAGTATGATGCCAGTGAAGAGACTTTTAAAGAGACCTCTAGAGAGTTACG 448

Qy 572 TTTCGGGACT--GCCCAAGTGGCCCTGATCCCGGCCCAACAGAGAGAGCGCTGGTGC 628
Db 449 TGAATAATTAAGTACCGTTACGTCATTCGATGAGCAGCGTTCTCGATTGATCA 508

Qy 629 GAGCCCGCTCTGGGGCGTCTCGCGGAGGGGCGATGTTCTGACCTTCTCGAGCTGTC 688
Db 509 GAGCTAGTTAAAGGTCTGCTGTGCTAAAGGCCAAGTATACCTTTTATAGACGGC 568

Qy 689 ACTGTGAGTCCCAAGAGGGTGTCTGAGCGCTGTCTGAGAGGATCCATGAAGAGAGT 748
Db 569 ACTGTGAGTGCACAGTGGGTGTCTCTCTAGCCAGATCAACATGACAGGA 628
Qy 749 CGGCACTGTGTGCCCGGTGATTGATGTGATCGACCTGGAAACACTTCGGAATACCTGGGA 808
Db 629 AGACAGTGGTCTGTCCCATCATAGATGTGATGACACTTTCGAGTACATGGCAG 688
Qy 809 ACTCCGGGAGCCCAAGATCGCGGTTCGACCTGGAGGCTGTGTTCACTGCGACACAG 868
Db 689 GTTCTGACATGACCTA---TGGCGGGTTCAACTGGAAGCTCAACTTTCGCTGGTATCCTG 745
Qy 869 TTCTCTGAGAGGGAGAG---ATACGGATGCAATCCCGCTCGATGTCTCATCAGTCTCCAA 925
Db 746 TTCCCAAGAGAAATGGACAGAGGAAGGTGATCGGACTCTCTCTGTGGAACACTTA 805
Qy 926 CAATGCTGTGTGGCTGTTTGTGTGAGTAAGAAATATTTTGAATATCTGGGTCTTATG 985
Db 806 CAATGCGAGGAGGCTTTTTCATAGACAGAGATTACTTTTCAGGAAATTTGGAACATATG 865
Qy 986 ATACAGGAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGT 1045
Db 866 ATCTGGAATGGATATTTTGGGAGGAGAAACCTAGAAATTTCTTTAGGATTTGGCAGT 925
Qy 1046 GTGTGGGGTCTTGGAAACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAG 1105
Db 926 GTGGAGAACTTTGGAGATGTTTACTTGTCTCAATGTTGACATGTTTCGGGAAAGCTA 985
Qy 1106 CTCCTACTCTCCGCAACAAAGGCTCTGGCCAAACAGTGT-----TGTGCGAG 1150
Db 986 CACCTTACAGTTTCCAGGAGGACGCGGCGAGATTATCAATAAAAAATAACAGACGACTTG 1045
Qy 1151 CTGAAGTATGGATGGATGAATTTAAAGAGCTCTTACTACCATCGCAACCCCGTCCCGCT 1210
Db 1046 CAGAAGTATGGATGGATGAATTTCAAGAAATTTCTTATATAATTTCTCCAGGTGTTACAA 1105
Qy 1211 TGGAACTTTTGGGATGTGACAGAGAGAAAGCAGCTCCGGGCAACAGCTCCAGTGTAAAG 1270
Db 1106 AGTAGATTATGGAGATATATCATCAAGACTTGGTCTAAGGCACAACTCCATGACAGAC 1165
Qy 1271 ACTTCAAGTGTCTTTGGAGACTGTGTATCCAGAACTGCGATGCTGCTGAGACAGGCGTG 1330
Db 1166 CATTTCTTGTGACTAGAGATATTTATCTGATTTCTCAGATTCCTCGTCACTATTCTTCT 1225
Qy 1331 GCTTCTCGGATGCTCCAGAAACAAAGGACTTAACAGACTACTGCTTTGACTATACCTTC 1390
Db 1226 CTTTGGGAGAGATAC-----GAAATGTGAAACAAATCAGTGTCTAGATAACATGGCTA 1279
Qy 1391 CGGATGAAACACAGATTGTGGGACACCAGGTCTCTGTACCTCTGTATGGGATGGCC 1450
Db 1280 GAAAAGAGAAATGAAAAGTTGGA-----ATTTTAACTGTCTATGGATGGAG 1327
Qy 1451 AGAATCAGTTTTCGAGTACACGTCGCCAGAAAGAAAT 1487
Db 1328 GTAATCAGGTTTCTCTTACACGTCGCCAAAGAAAT 1364

RESULT 10

US-08-967-508-11
; Sequence 11, Application US/08967508
; Patent No. 5910570
; GENERAL INFORMATION:

; APPLICANT: The Upjohn Company
; APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNac:
; TITLE OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual
; ADDRESSEE: Property Legal Services
; STREET: 301 Henrietta Street
; CITY: Kalamazoo

: REFERENCE/DOCKET NUMBER: 4755.P CP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 616-833-2211

TELEFAX: 616-833-8897

TELEX: 224401

; INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 1680 base pairs

```

;
; TYPE: nucleic acid

```

; STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-967-506-11

Query Match 13.3%; Score 232.2; DB 3; Length 1680;

Best Local Similarity 54.8%; Pred. No. 5.3e-44;

Matches 612; Conservative 0; Mismatches 463; Indels 42; Gaps 6;

392 ATTATGATAATTGCCCCAGGACATCTGTTATCATAGCATTTTATATAATGAAGCCTGGTCAA 451

DD 332 ATCCAGATATACCTTCTTACACACAGATGTGTGTATGTATTTTCCACACATAGAGGCTTCTGACGCA, 333

452 CTCTCCTTCGGACAGTTACAGTGTCTTGAGACATCCCCGGATATCCTGCTAGAAG 511

Db 392 CACTTCTGCCGAACGTGTCCATAGCGTCATTAAATCGCTCACCAAGGCACATGCTAGAAGAA 451

THE UNIVERSITY OF CHICAGO

Qy

Db 452 TTGTTCTAGTAGATGCCAGTGAAGAGACCTTTTAAAAAGACCTCTAGAGGTTACG 511

572 TTTCCGGGACT--GCCCAAGGTGCGCCCTGATCCGGCGCCAACAAGAGAGAGGGCCTGGTGC 628

Qy

DB 312 TGAATAAATTAATAAGTACCCGTTACGGTCAATCGATGGAGCAGCGTTCGGATTATCA 371

629 GAGCCCGGTCCTGGGGCGTCCTGCGGCGAGGGGGGATGTTCTGACCTTCCTGGACTGTC 688

db 572 GAGCTAGGTTAAAAGGTGCTGCTGTCTAAAGGCCAAGTGATCATCATTTTATAGACGGC 631

DB
372 GAGCTAGGTTAATAAGGGTGTGTGTATCAACCTTTTTTAGAGCGCC

QY 689 AC TGT GAG TGC CAC GAAG GGT TGG CTGG AGC CGT GCT GCA GAG GAT CCA TGA GAG GAG T / 48

d_b 632 A C T G T G A G T G C A C A G T G G G T G G C T C T C T A G C C A G G A T C A A C A T G A C A G G A 691

ON 749 CCGCAGTCCGCTGTGCCCGGTGATTGATGTGATCGACTTGGAAACACCTTCGAATACCTGGGGA 808

QY 749 CGGCAGTGGTGTGCTCCGGTGATGATGIGATCGACTCGAAACACCTTCGAATACCTGGGA 808

DB 692 AGACAGTGGTCTGTCTCCCATCATAGATGTGATCAGTGAIGACACTTTCGAGTACATGGCAG 751

808 A T C C C G G A G C C C C A G A T C G G C G G T T T C G A C T C G A G G C T G T G T T C A C G T G G C A C A C A G 868

753 CTTCTCAGCATCCCTTCAATCGACGCTCAATTTCGTGGTATCTCTG 808

DB 752 GTTC TGACA TGACCTA --- TGGCGGGTTC AAC TGGGAAGCTCAACTTTCGCTGGGATCCCG 808

QY
869 TTTCTGAGAGGGAGAGG---ATACGGATGCAATCCCGTCTGATGTCATCAGGTCCTCCAA 923

909 TTCCCCCAACACCAATCCCAACCAACGTCATCCGATCCGACACCTA 868

5 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1

QY 926 CAATGGCTGGTGGGCTGTTTGCTGTGAGTAAGAAATAATTTGAATAATC"GGGGCT"TAG 985

Db 869 CAATGGCAGGAGGCCCTTTTTCATAGACACAGAGATTACTTTCAGGAATTTGGAACATATG 928

1045

Db 1281 CATTCTCTTGGTACCTAGAGAATATTTATCCTGATTTCTCAGATTCTCGTCACTATTCT 1340
Qy 1331 GCTTCTTCGGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGACTATTAACCTC 1390
Db 1341 CTTTGGGAGAGATAC-----GAAATGTGGAAACAAATCAGTGTCTAGATAACATGGCTA 1394
Qy 1391 CCGATGAAACACAGATTGTGGGACACCAGGTCATTCTGTACTCTGTCAATGGGATGGCC 1450
Db 1395 GAAAGAGATGAAAGTTGA-----ATTTTAACTGTCAATGGTATGGAG 1442
Qy 1451 AGAATCAGTTTTTCGAGTACACGTCCTCCAGAAAGAAAT 1487
Db 1443 GTAATCAGTTTTTCTTTACACTGCTCAACAAAGAAAT 1479

Search completed: December 7, 2003, 21:21:39
Job time : 109.153 secs

Qy 121 CGTGGGCGGGCGGGCTGCGAGCCGGGACCCCGCGCACCCCGCGCGCGCGCGGCGG 180
Db 121 CGTGGGCGGGCGGGCTGCGAGCCGGGACCCCGCGCACCCCGCGCGCGCGCGGCGG 180
Qy 181 GCGAGCGGTCATGCGG 240
Db 181 GCGAGCGGTCATGCGG 240
Qy 241 GCGTGGCGCTGCGAGTGCAGGGCGAGGAGCTGCGGCTGCGAGGAGGAGCGTGGCGCTG 300
Db 241 GCGTGGCGCTGCGAGTGCAGGGCGAGGAGCTGCGGCTGCGAGGAGGAGCGTGGCGCTG 300
Qy 301 CACGAGATTAACATCTACCTCAGGACCGCATCTCACTGCAACCGCGCTGCCCGAGCGC 360
Db 301 CACGAGATTAACATCTACCTCAGGACCGCATCTCACTGCAACCGCGCTGCCCGAGCGC 360
Qy 361 TGGAAACCGCTGTGCAAGAGAGAAATATGATTAATGATTAATTTGCCAGGACATCTGTT 420
Db 361 TGGAAACCGCTGTGCAAGAGAGAAATATGATTAATGATTAATTTGCCAGGACATCTGTT 420
Qy 421 ATCATAGCATTTTATATGAGGCTGCTCAACTCTCTTCCGACAGTTTACAGTGTCTT 480
Db 421 ATCATAGCATTTTATATGAGGCTGCTCAACTCTCTTCCGACAGTTTACAGTGTCTT 480
Qy 481 GAGACATCCCGGATATCTCTAGAGAGAGTGTCTTGTAGATGACTACAGTGTATAGA 540
Db 481 GAGACATCCCGGATATCTCTAGAGAGAGTGTCTTGTAGATGACTACAGTGTATAGA 540
Qy 541 GAGCACTGGAAGAGCGCTTGGCCAAATGAGCTTTTGGGACTGCCCCAAGGTCGCTGATC 600
Db 541 GAGCACTGGAAGAGCGCTTGGCCAAATGAGCTTTTGGGACTGCCCCAAGGTCGCTGATC 600
Qy 601 CGCGCAACAGAGAGAGGCGCTGTCGAGCGCGCTGTCGGGGCGCTGCGCGCGAGG 660
Db 601 CGCGCAACAGAGAGAGGCGCTGTCGAGCGCGCTGTCGGGGCGCTGTCGGCGCGAGG 660
Qy 661 GCGGATGTTCTGACCTTCTGAGCTGTCACTGTGAGTGCCACGAAGGTCGCTGAGCGG 720
Db 661 GCGGATGTTCTGACCTTCTGAGCTGTCACTGTGAGTGCCACGAAGGTCGCTGAGCGG 720
Qy 721 CTGCTGAGAGGATCCATGAAGAGAGTTCGCGAGTGTGTCGCCGCTGATGATGATG 780
Db 721 CTGCTGAGAGGATCCATGAAGAGAGTTCGCGAGTGTGTCGCCGCTGATGATGATG 780
Qy 781 GACTGGAACCTTCGATACCTCGGNACTCCCGGAGCCCGAGTCCGAGTTCGCGGTTCCG 840
Db 781 GACTGGAACCTTCGATACCTCGGNACTCCCGGAGCCCGAGTCCGAGTTCGCGGTTCCG 840
Qy 841 TGGAGGCTGTGTTTCACTGCGCACACAGTTCCTGAGAGGAGAGATACGATGCAATCC 900
Db 841 TGGAGGCTGTGTTTCACTGCGCACACAGTTCCTGAGAGGAGAGATACGATGCAATCC 900
Qy 901 CCGGTCGATGTCATCAGTCTCCAAATAGCTGCTGGGCTGTTTGTGTGAGTAAGAAA 960
Db 901 CCGGTCGATGTCATCAGTCTCCAAATAGCTGCTGGGCTGTTTGTGTGAGTAAGAAA 960
Qy 961 TATTTGATATCTGGGCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTC 1020
Db 961 TATTTGATATCTGGGCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAACCTC 1020
Qy 1021 GAAATTTCTTTAGGATCTGCGAGTGTGCTGGGTTCTGGAACACACCCATGTTCCCAT 1080
Db 1021 GAAATTTCTTTAGGATCTGCGAGTGTGCTGGGTTCTGGAACACACCCATGTTCCCAT 1080
Qy 1081 GTTGCCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAAGGCTCTGCGCAACAGT 1140
Db 1081 GTTGCCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAAGGCTCTGCGCAACAGT 1140
Qy 1141 GTTGTGAGCTGAAGTATGATGATGAATTAAGAGCTCTACTACCATCGCAACCCC 1200
Db 1141 GTTGTGAGCTGAAGTATGATGATGAATTAAGAGCTCTACTACCATCGCAACCCC 1200
Qy 1201 CGTGGCGCTTGGAACTTTTGGGATGTGACAGAGAGAGAGCTCCCGGACAAAGCTC 1260

Db 1201 CGTGGCGCTTGGAACTTTTGGGATGTGACAGAGAGAGAGCTCCCGGACAAAGCTC 1260
Qy 1261 CAGTGTAAAGACTTCAAGTGGTTCTTGGAGACTGTGTATCCAGAACTGCACTGTGCCCTGAG 1320
Db 1261 CAGTGTAAAGACTTCAAGTGGTTCTTGGAGACTGTGTATCCAGAACTGCACTGTGCCCTGAG 1320
Qy 1321 GACAGGCTCGCTTCTTCCGGATGTCCAGAAACAAAGGACTAAACAGACTACTGTCTTTGAC 1380
Db 1321 GACAGGCTCGCTTCTTCCGGATGTCCAGAAACAAAGGACTAAACAGACTACTGTCTTTGAC 1380
Qy 1381 TATAACCTCCCGATGAAAAACCAAGATTGTGGGACACAGGTCACTTGTGTACCTGTGTCAT 1440
Db 1381 TATAACCTCCCGATGAAAAACCAAGATTGTGGGACACAGGTCACTTGTGTACCTGTGTCAT 1440
Qy 1441 GGGATGGCCAGAACTAGTTTTTTCGAGTACACGCTCCAGAAAGAAATACGCTATAACACC 1500
Db 1441 GGGATGGCCAGAACTAGTTTTTTCGAGTACACGCTCCAGAAAGAAATACGCTATAACACC 1500
Qy 1501 CACGAGCTCAGGCTGCTGCTTGTGGAAGCAGGAATGATACCTTATCATGCACTCTC 1560
Db 1501 CACGAGCTCAGGCTGCTGCTTGTGGAAGCAGGAATGATACCTTATCATGCACTCTC 1560
Qy 1561 TGGGAAGAACTCTCCCGAGAGAAATCAGAACTTCACTTTCAGGAGGATGGAATCTTATTT 1620
Db 1561 TGGGAAGAACTCTCCCGAGAGAAATCAGAACTTCACTTTCAGGAGGATGGAATCTTATTT 1620
Qy 1621 CAGCAAGCTCCAAAGAAATGTGTCCAGGCTGCGAGGAGGATGCGAGTTCGTT 1680
Db 1621 CAGCAAGCTCCAAAGAAATGTGTCCAGGCTGCGAGGAGGATGCGAGTTCGTT 1680
Qy 1681 CCCTCTTACGAGACTGCACCACTCGGATCATCAGAAATGGTCTTCAAAGAGCGCATG 1740
Db 1681 CCCTCTTACGAGACTGCACCACTCGGATCATCAGAAATGGTCTTCAAAGAGCGCATG 1740
Qy 1741 TTATGA 1746
Db 1741 TTATGA 1746

RESULT 2
US-10-074-527-1
; Sequence 1, Application US/10074527
; Publication No. US20020142426A1
; GENERAL INFORMATION:
; APPLICANT: Olandt, Peter J.
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Galvin, Katherine A.
; APPLICANT: Millennium Pharmaceuticals Inc.
; TITLE OF INVENTION: 33945, A Human Glycosyltransferase and
; FILE REFERENCE: MPI2001-018P1(M)
; CURRENT APPLICATION NUMBER: US/10/074,527
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/269202
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2850
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (81)...(1826)
US-10-074-527-1

Query Match 100.0%; Score 1746; DB 13; Length 2850;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1746; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGTGGGGGCGCAGCGCGCGCGCTGCGCGGAACTGCGCGCGCGCGCGCGCGCGCGG 60

Db 81 ATGTGGGGCGCAGCGGCGGCGCTGCCCCGGGAACTGCGGCGCGCGCGGAGCG 140
Qy 61 CTGTGTGTCTCTGGCGCTACTGGCGTTGGCGGGCTGGGCTGGTCTGGCGGCGCAG 120
Db 141 CTGTGTGTCTCTGGCGCTACTGGCGTTGGCGGGCTGGGCTGGTCTGGCGGCGCAG 200
Qy 121 CGTGGGGCGGGGCGGGGCTGCGGAGCGGGACCCCGCGCACCCCGCGCCCCCGGGCG 180
Db 201 CGTGGGGCGGGGCGGGGCTGCGGAGCGGGACCCCGCGCACCCCGCGCCCCCGGGCG 260
Qy 181 CGGAGCGCGTCTATGCGCGCGCGCGCTGCGGCGAAACGCGTGGCGCGCGGGGCGAG 240
Db 261 CGGAGCGCGTCTATGCGCGCGCGCGCTGCGGCGAAACGCGTGGCGCGCGGGGCGAG 320
Qy 241 GCGGTGCGGTGTCAGCTGTCAGGGCGAGAGCTGCGGCTGCGAGGAGAGAGCTGCGGCTG 300
Db 321 GCGGTGCGGTGTCAGCTGTCAGGGCGAGAGCTGCGGCTGCGAGGAGAGAGCTGCGGCTG 380
Qy 301 CACGAGATTAACTACTACTCAGGACCGCATCTCACTGACCGCGCGCTGCCGAGCGC 360
Db 381 CACGAGATTAACTACTACTCAGGACCGCATCTCACTGACCGCGCGCTGCCGAGCGC 440
Qy 361 TGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATAATTGCCCCAGGACATCTGTT 420
Db 441 TGGAAACCGCTGTGCAAGAGAGAAATATGATTATGATAATTGCCCCAGGACATCTGTT 500
Qy 421 ATCATAGCATTTTATAATGAAGCTGTGTCACCTCTCTCCGAGAGTTTACAGTGTCTTT 480
Db 501 ATCATAGCATTTTATAATGAAGCTGTGTCACCTCTCTCCGAGAGTTTACAGTGTCTTT 560
Qy 481 GAGACATCCCGGATATCTCTGATAGAGAGTGAATCTTGTAGTGAATCACTGAGTATGATG 540
Db 561 GAGACATCCCGGATATCTCTGATAGAGAGTGAATCTTGTAGTGAATCACTGAGTATGATG 620
Qy 541 GAGCACCCTGAAGAGAGCGCTTGGCCAAATGAGCTTTGGGCACTGCCAAGGTGGCGCTGATC 600
Db 621 GAGCACCCTGAAGAGAGCGCTTGGCCAAATGAGCTTTGGGCACTGCCAAGGTGGCGCTGATC 680
Qy 601 CGCGCCAAAGAGAGAGCGCTTGGTGCAGCGCGGCTGTGGGGGCGCTGTGGCGGAGG 660
Db 681 CGCGCCAAAGAGAGAGCGCTTGGTGCAGCGCGGCTGTGGGGGCGCTGTGGCGGAGG 740
Qy 661 GCGGATGTTCTGACCTTCTGAGTGTGACCTGTGAGTGTGACGAGAGGAGTGTGAGAGCG 720
Db 741 GCGGATGTTCTGACCTTCTGAGTGTGACCTGTGAGTGTGACGAGAGGAGTGTGAGAGCG 800
Qy 721 CTGCTGCAGAGGATCCATGAAGAGGAGTGGGCACTGGGCTGGCGGTGATTGATGTGATC 780
Db 801 CTGCTGCAGAGGATCCATGAAGAGGAGTGGGCACTGGGCTGGCGGTGATTGATGTGATC 860
Qy 781 GACTGGAACACCTTTCGAATACCTTGGGAACTCCGGGAGCCCGAGATCGCGGTTTCGAC 840
Db 861 GACTGGAACACCTTTCGAATACCTTGGGAACTCCGGGAGCCCGAGATCGCGGTTTCGAC 920
Qy 841 TGAGGCTGTGTTCAGTGGGCACACAGTTCTTGAGAGGAGAGGATACGGATGCAATCC 900
Db 921 TGAGGCTGTGTTCAGTGGGCACACAGTTCTTGAGAGGAGAGGATACGGATGCAATCC 980
Qy 901 CCGGTGATGTCTATCAGTCTCAACATCGCTGTGGGCTGTTGCTGTGAGTGAAGAA 960
Db 981 CCGGTGATGTCTATCAGTCTCAACATCGCTGTGGGCTGTTGCTGTGAGTGAAGAA 1040
Qy 961 TATTTTGAATATCTGGGCTTATGATACAGGAATGGAAGTTGGGAGGAGAGAAACCTC 1020
Db 1041 TATTTTGAATATCTGGGCTTATGATACAGGAATGGAAGTTGGGAGGAGAGAAACCTC 1100
Qy 1021 GAATTTTCTTTAGGATCTGGCAGTGTGGGTTCTGGAACACACCCATGTTCCTCAT 1080
Db 1101 GAATTTTCTTTAGGATCTGGCAGTGTGGGTTCTGGAACACACCCATGTTCCTCAT 1160
Qy 1081 GTTGGCCATGTTTTCCCGAGAGCTCCCTACTCCCGGCAACAGGCTCTGGCCACAGT 1140
Db 1161 GTTGGCCATGTTTTCCCGAGAGCTCCCTACTCCCGGCAACAGGCTCTGGCCACAGT 1220

Qy 1141 GTTCTGTCAGCTCAAGTATGATGGATGAATTTTAAAGAGCTCTACTACCATCGCAACCCC 1200
Db 1221 GTTCTGTCAGCTCAAGTATGATGGATGAATTTTAAAGAGCTCTACTACCATCGCAACCCC 1280
Qy 1201 CGTCCCGCTTGGAACTTTTGGGATGTGACAGAGGAGGAGAGCTCCGGGACAGCTC 1260
Db 1281 CGTCCCGCTTGGAACTTTTGGGATGTGACAGAGGAGGAGAGCTCCGGGACAGCTC 1340
Qy 1261 CAGTGTAAAGACTTCAAGTGTGTTTGGGAGCTGTGTATCCAGAACTGCAATGTGCTTGG 1320
Db 1341 CAGTGTAAAGACTTCAAGTGTGTTTGGGAGCTGTGTATCCAGAACTGCAATGTGCTTGG 1400
Qy 1321 GACAGGCTGGCTTCTTCCGGATGCTCCAGAAACAAAGGACTAAACAGACTACTGCTTTGAC 1380
Db 1401 GACAGGCTGGCTTCTTCCGGATGCTCCAGAAACAAAGGACTAAACAGACTACTGCTTTGAC 1460
Qy 1381 TATAACCTCCCGATGAAACACAGATTGTTGGGACACAGCTCATCTGTACTCTGTCTAT 1440
Db 1461 TATAACCTCCCGATGAAACACAGATTGTTGGGACACAGCTCATCTGTACTCTGTCTAT 1520
Qy 1441 GGGATGGGCGCAGAACTCAGTTTTCGAGTACACGTCGAGAAAGAAATACGCTATAACACC 1500
Db 1521 GGGATGGGCGCAGAACTCAGTTTTCGAGTACACGTCGAGAAAGAAATACGCTATAACACC 1580
Qy 1501 CACGAGCTGAGGGCTGCATTGCTGTGGAAGCAGGAATGGATACCCCTTATCATGCTCTC 1560
Db 1581 CACGAGCTGAGGGCTGCATTGCTGTGGAAGCAGGAATGGATACCCCTTATCATGCTCTC 1640
Qy 1561 TCGGAAGAACTCCCGCAGAGAACTCAGAGTTTCACTCTGCGAGGAGTGGATCTTTATTT 1620
Db 1641 TCGGAAGAACTCCCGCAGAGAACTCAGAGTTTCACTCTGCGAGGAGTGGATCTTTATTT 1700
Qy 1621 CAGGAAACAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAGGAGTGGATCTTTATTT 1680
Db 1701 CAGGAAACAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAGGAGTGGATCTTTATTT 1760
Qy 1681 CCATCTTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1740
Db 1761 CCATCTTACGAGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1820
Qy 1741 TTATGA 1746
Db 1821 TTATGA 1826

RESULT 3

US-10-198-846-9749/c
; Sequence 9749, Application US/10198846
; Publication No. US2003099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9749
; LENGTH: 2904
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1, 2, 3, 4, 5, 2903, 2904
; OTHER INFORMATION: n = A,T,C or G

US-10-198-846-9749

Query Match		100.0%;	Score 1746;	DB 14;	Length 2904;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches: 1746;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGTGGGGGCGCACGGCGCGGCGCTGCGCCGCGGAACCTGCGGGCGCGCCCGGAGGCG	60		
Db	2826	ATGTGGGGGCGCACGGCGCGGCGCTGCGCCGCGGAACCTGCGGGCGCGCCCGGAGGCG	2767		
QY	61	CTGTTGGTGTCTCTGGCGCTACTGCGGTTTGGCCGCGCTGGGCTCGTGCTGCGGGCGCAG	120		
Db	2766	CTGTTGGTGTCTCTGGCGCTACTGCGGTTTGGCCGCGCTGGGCTCGTGCTGCGGGCGCAG	2707		
QY	121	CGTGGGGCGGGGCGGGGCTGCGAGCGGGACCCCGCGCACCCCGCGCCCGCGGCGG	180		
Db	2706	CGTGGGGCGGGGCGGGGCTGCGAGCGGGACCCCGCGCACCCCGCGCCCGCGGCGG	2647		
QY	181	CGCGAGCGGTCTATGCGCGCGCGCGCTGCGCGCGGAACGCGCTGGGCGCGCGGGCGAG	240		
Db	2646	CGCGAGCGGTCTATGCGCGCGCGCGCTGCGCGCGGAACGCGCTGGGCGCGCGGGCGAG	2587		
QY	241	GCGGTGGGCTGCAGCTGCGAGGGCGAGGAGCTGCGGCTGCAGAGAGAGAGCTGCGGCTG	300		
Db	2586	GCGGTGGGCTGCAGCTGCGAGGGCGAGGAGCTGCGGCTGCAGAGAGAGAGCTGCGGCTG	2527		
QY	301	CACAGATTAAACATCTACCTCAGCGACCGCATCTCAGTGCACCGCGCCCTGCCGAGCGC	360		
Db	2526	CACAGATTAAACATCTACCTCAGCGACCGCATCTCAGTGCACCGCGCCCTGCCGAGCGC	2467		
QY	361	TGGAACCGCTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCAGGACATCTGTT	420		
Db	2466	TGGAACCGCTGTGCAAGAGAGAAATATGATTATGATTAATTTGCCAGGACATCTGTT	2407		
QY	421	ATCATAGCATTTTATAATGAAGCTGTGTCAACTCTCTTCGGACAGTTTACAGTGCCTT	480		
Db	2406	ATCATAGCATTTTATAATGAAGCTGTGTCAACTCTCTTCGGACAGTTTACAGTGCCTT	2347		
QY	481	GAGACATCCCGGATATCTCTAGAGAAAGTATCTCTGTAGATGACTACAGTGATAGA	540		
Db	2346	GAGACATCCCGGATATCTCTAGAGAAAGTATCTCTGTAGATGACTACAGTGATAGA	2287		
QY	541	GAGCACTGAAGGAGCGCTTGGCCAAATGAGCTTTTGGGACTGCGCAAGGTGCGCTGATC	600		
Db	2286	GAGCACCTGAAGGAGCGCTTGGCCAAATGAGCTTTTGGGACTGCGCAAGGTGCGCTGATC	2227		
QY	601	CGCGCCAAAGAGAGAGGCGCTGTGTCGAGCCGCGCTGTCTGGGGCGCTCTGCGGCGAGG	660		
Db	2226	CGCGCCAAAGAGAGAGGCGCTGTGTCGAGCCGCGCTGTCTGGGGCGCTCTGCGGCGAGG	2167		
QY	661	GGCGATGTTCTGACCTTCTTGACTGTGTCATCTGTAGTGCACGAAAGGTTGGCTGGAGCCG	720		
Db	2166	GGCGATGTTCTGACCTTCTTGACTGTGTCATCTGTAGTGCACGAAAGGTTGGCTGGAGCCG	2107		
QY	721	CTGCTGCAGAGGATCCATGAAGAGGAGTGGCAGTGTGTGCCGGTGATGTGATGTCATC	780		
Db	2106	CTGCTGCAGAGGATCCATGAAGAGGAGTGGCAGTGTGTGCCGGTGATGTGATGTCATC	2047		
QY	781	GACTGGAAACCTTTCGAATACCTCGGGGAACTCCGGGAGCGCCACAGATCGCGGTTTCGAC	840		
Db	2046	GACTGGAAACCTTTCGAATACCTCGGGGAACTCCGGGAGCGCCACAGATCGCGGTTTCGAC	1987		
QY	841	TGGAGGCTGTGTTCACGTGGCAACAGTTCTCTGAGAGGGAGAGGATACGGATGCAATCC	900		
Db	1986	TGGAGGCTGTGTTCACGTGGCAACAGTTCTCTGAGAGGGAGAGGATACGGATGCAATCC	1927		
QY	901	CCGTCGATGTCATCAGGTCTCCAAATGCTGCTGGGCTGTTTGTGTGAGTAAGAAA	960		
Db	1926	CCGTCGATGTCATCAGGTCTCCAAATGCTGCTGGGCTGTTTGTGTGAGTAAGAAA	1867		
QY	961	TATTTTGAATATCTGGGGTCTTATGATACAGAAATGGAAGTTTGGGAGGAGAAAACTC	1020		
Db	1866	TATTTTGAATATCTGGGGTCTTATGATACAGAAATGGAAGTTTGGGAGGAGAAAACTC	1807		

Qy	1021	GAATTTTCTTTAGGATCTGGCAGTCTGGTGGGTTCTGGAAACACACCCATGTTCCCAT	1080
Db	1806	GAATTTTCTTTAGGATCTGGCAGTCTGGTGGGTTCTGGAAACACACCCATGTTCCCAT	1747
Qy	1081	GTTGGGCATGTTTCCCAAGCAGCTCCCTACTCCGCAACAAGGCTCTGGGCAACAGT	1140
Db	1746	GTTGGGCATGTTTCCCAAGCAGCTCCCTACTCCGCAACAAGGCTCTGGGCAACAGT	1687
Qy	1141	GTTTCGTGAGCTGAGTATGGATGAATTTAAAGAGCTCTACTACCATCCCAACCCC	1200
Db	1686	GTTTCGTGAGCTGAGTATGGATGAATTTAAAGAGCTCTACTACCATCCCAACCCC	1627
Qy	1201	CGTGCCTGCTTGGAACTTTTGGGGATGTGACAGAGAGGAGCAGCTCCGGGCAAGCTC	1260
Db	1626	CGTGCCTGCTTGGAACTTTTGGGGATGTGACAGAGAGGAGCAGCTCCGGGCAAGCTC	1567
Qy	1261	CAGTGTAAAGACTTCAAGTGGTCTTGGGAGACTGTGTATCCAGAACTGCAATGTCCTGAG	1320
Db	1566	CAGTGTAAAGACTTCAAGTGGTCTTGGGAGACTGTGTATCCAGAACTGCAATGTCCTGAG	1507
Qy	1321	GACAGGCTGGCTCTTCCGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGAC	1380
Db	1506	GACAGGCTGGCTCTTCCGGATGCTCCAGAACAAAGGACTAACAGACTACTGCTTTGAC	1447
Qy	1381	TATAACCTCCCGATGAAAAACCAAGATTGTGGGACACACAGGTCTTCTGTACCTCTGTAT	1440
Db	1446	TATAACCTCCCGATGAAAAACCAAGATTGTGGGACACACAGGTCTTCTGTACCTCTGTAT	1387
Qy	1441	GGATGGGCGCAGAAATCAGTTTTTCAGTATACACTCCAGAAAGAAATACCGTATACACC	1500
Db	1386	GGATGGGCGCAGAAATCAGTTTTTCAGTATACACTCCAGAAAGAAATACCGTATACACC	1327
Qy	1501	CACAGGCTGAGGCTGCAATGCTGTGGAGAGGAGTGAATACCTTATCATGATCTC	1560
Db	1326	CACAGGCTGAGGCTGCAATGCTGTGGAGAGGAGTGAATACCTTATCATGATCTC	1267
Qy	1561	TGCGAAGAACTGCCCGCAGAGAAATCAGAGTTTCTTTCAGGAGGAGTGGATCTTTATTT	1620
Db	1266	TGCGAAGAACTGCCCGCAGAGAAATCAGAGTTTCTTTCAGGAGGAGTGGATCTTTATTT	1207
Qy	1621	CACGAAACAGTCCCAAGAAATGTGTCAGGCTGCGAGGAGGAGTGCAGTTCGTT	1680
Db	1206	CACGAAACAGTCCCAAGAAATGTGTCAGGCTGCGAGGAGGAGTGCAGTTCGTT	1147
Qy	1681	CCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATG	1740
Db	1146	CCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCATG	1087
Qy	1741	TTATGA 1746	
Db	1086	TTATGA 1081	

RESULT 4
US-10-292-896-57
; Sequence 57, Application US/10292896
; Publication No. US20030186850A1
; GENERAL INFORMATION:
; APPLICANT: HASSAN, Helle
; APPLICANT: REIS, Celso A.
; APPLICANT: BENNETT, Eric P.
; APPLICANT: CLAUSEN, Henrik
; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GA
; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS
; FILE REFERENCE: 4305/1H154-US3
; CURRENT APPLICATION NUMBER: US/10/292,896
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/425,204
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/DK01/00328
; PRIOR FILING DATE: 2001-05-10

; PRIOR APPLICATION NUMBER: US 60/203,331
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
; LENGTH: 1746
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-292-896-57

Query Match 99.9%; Score 1744.4; DB 12; Length 1746;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1745; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGTGGGGCCACGCGCGCGCGCTGCGCGCGGAACTGCGCGCGCGCGCGCGGAGCG 60
Db 1 ATGTGGGGCCACGCGCGCGCGCGCTGCGCGCGGAACTGCGCGCGCGCGCGGAGCG 60
Qy 61 CTGTTGATGCTCTGGCGCTACTGGCGTTGGCCGGGCTGGGCTGGTCTGGCGGCGAG 120
Db 61 CTGTTGATGCTCTGGCGCTACTGGCGTTGGCCGGGCTGGGCTGGTCTGGCGGCGAG 120
Qy 121 CGTGGGCGCGGGCGCGGGCTGCGAGCGGGACCCCGCGCACCCCGCGCGCGCGGCGG 180
Db 121 CGTGGGCGCGGGCGCGGGCTGCGAGCGGGACCCCGCGCACCCCGCGCGCGCGGCGG 180
Qy 181 CGGAGCGGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 240
Db 181 CGGAGCGGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 240
Qy 241 GCGGTGGGCTGAGCTGCGAGGCGAGAGTGGCGCTGCGAGGAGAGAGTGGCGCTG 300
Db 241 GCGGTGGGCTGAGCTGCGAGGCGAGAGTGGCGCTGCGAGGAGAGAGTGGCGCTG 300
Qy 301 CACGAGATTAACATCTACCTCAGGACCGCATCTCACTGCGACCGCGCGCGCGCGCGG 360
Db 301 CACGAGATTAACATCTACCTCAGGACCGCATCTCACTGCGACCGCGCGCGCGCGG 360
Qy 361 TGGAAACCGCTGTGCAAGAGAAATATGATTAATTAATTTGCCAGGACATCTGTT 420
Db 361 TGGAAACCGCTGTGCAAGAGAAATATGATTAATTAATTTGCCAGGACATCTGTT 420
Qy 421 ATCATAGCATTTTATATGAAGCGTGGTCAACTCTCTCTCGGACAGTTTACAGTGCCTT 480
Db 421 ATCATAGCATTTTATATGAAGCGTGGTCAACTCTCTCTCGGACAGTTTACAGTGCCTT 480
Qy 481 GAGACATCCCGGATATCTCTGCTAGAGAGAGTGTCTGTAGATGACTACAGTGATGA 540
Db 481 GAGACATCCCGGATATCTCTGCTAGAGAGAGTGTCTGTAGATGACTACAGTGATGA 540
Qy 541 GAGCACCCTGAAGGAGCGCTTGGCCAAATGAGCTTTTCGGGACTGCCCAAGGTGCGCTGATC 600
Db 541 GAGCACCCTGAAGGAGCGCTTGGCCAAATGAGCTTTTCGGGACTGCCCAAGGTGCGCTGATC 600
Qy 601 CGGCGCAACAGAGAGAGGCGCTGGTGGAGCGCGCGCTGCTGGGGCGTCTGGCGGAGG 660
Db 601 CGGCGCAACAGAGAGAGGCGCTGGTGGAGCGCGCGCTGCTGGGGCGTCTGGCGGAGG 660
Qy 661 GGGATGCTCTGACCTTCTGACCTGCTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAG 720
Db 661 GGGATGCTCTGACCTTCTGACCTGCTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAG 720
Qy 721 CTGCTGAGAGGATCCATGAAGAGAGTGGGAGTGGTGGCGGCTGGTGTGATGTGATC 780
Db 721 CTGCTGAGAGGATCCATGAAGAGAGTGGGAGTGGTGGCGGCTGGTGTGATGTGATC 780
Qy 781 GACTGGAAACCTTTCGAATACCTGGGAACTCCGGGAGCGCCAGATCGGCGGTTTCGAC 840
Db 781 GACTGGAAACCTTTCGAATACCTGGGAACTCCGGGAGCGCCAGATCGGCGGTTTCGAC 840
Qy 841 TGGAGGCTGTGTTTACGTGGCACACAGTTCCTGAGAGGAGGAGGATACGGATGCAATCC 900
Db 841 TGGAGGCTGTGTTTACGTGGCACACAGTTCCTGAGAGGAGGAGGATACGGATGCAATCC 900

RESULT 5
US-09-925-299-219
; Sequence 219, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883

Qy 901 CCCGTGATGTCATCAGGCTCTCAACAAATGGCTGGTGGGCTGTTGTGTGAGTAAGAAA 960
Db 901 CCCGTGATGTCATCAGGCTCTCAACAAATGGCTGGTGGGCTGTTGTGTGAGTAAGAAA 960
Qy 961 TATTTGATATCTGGGCTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAAACCTC 1020
Db 961 TATTTGATATCTGGGCTCTTATGATACAGGAATGGAAGTTTGGGAGGAGAAAACCTC 1020
Qy 1021 GAATTTTCCCTTTAGGATCTGGCAGTGTGGGCTCTGGAACACACACCCATGTTCCCAT 1080
Db 1021 GAATTTTCCCTTTAGGATCTGGCAGTGTGGGCTCTGGAACACACACCCATGTTCCCAT 1080
Qy 1081 GTTGGCCATGTTTCCCAAGAGCTCCCTACTCCCGCAACAGGCTCTGGCCCAACAGT 1140
Db 1081 GTTGGCCATGTTTCCCAAGAGCTCCCTACTCCCGCAACAGGCTCTGGCCCAACAGT 1140
Qy 1141 GTTGGCCATGTTTCCCAAGAGCTCCCTACTCCCGCAACAGGCTCTGGCCCAACAGT 1200
Db 1141 GTTGGCCATGTTTCCCAAGAGCTCCCTACTCCCGCAACAGGCTCTGGCCCAACAGT 1200
Qy 1201 CGTGGCCGCTTGGAACTTTTGGGATGTGACAGAGAGGAGCAGCTCCGGGCAACGCTC 1260
Db 1201 CGTGGCCGCTTGGAACTTTTGGGATGTGACAGAGAGGAGCAGCTCCGGGCAACGCTC 1260
Qy 1261 CAGTGTAAAGACTTCAAGTGGTCTTGGGAGCTGTGTATCCAGAACTGTCATGTCCTGAG 1320
Db 1261 CAGTGTAAAGACTTCAAGTGGTCTTGGGAGCTGTGTATCCAGAACTGTCATGTCCTGAG 1320
Qy 1321 GACAGGCTGGCTCTTCCGGATGCTCCAGAAACAAAGGACTAACAGACTACTGCTTGAC 1380
Db 1321 GACAGGCTGGCTCTTCCGGATGCTCCAGAAACAAAGGACTAACAGACTACTGCTTGAC 1380
Qy 1381 TATAACCTCCCGATGAAAACCAAGATTGCGGACACAGCTCATCTGTACCTCTGTCAT 1440
Db 1381 TATAACCTCCCGATGAAAACCAAGATTGCGGACACAGCTCATCTGTACCTCTGTCAT 1440
Qy 1441 GGGATGGCCAGAAATCAGTTTTTTCAGTACACGCTCCAGAAAGAAATACGCTATAACACC 1500
Db 1441 GGGATGGCCAGAAATCAGTTTTTTCAGTACACGCTCCAGAAAGAAATACGCTATAACACC 1500
Qy 1501 CACGAGCTGAGGCTGCTATGCTGTGGAAGAGAGGATGATGATGATGATGATGATGATC 1560
Db 1501 CACGAGCTGAGGCTGCTATGCTGTGGAAGAGAGGATGATGATGATGATGATGATC 1560
Qy 1561 TCGAAGAAACTGCCCGCAGAGATCAGAGTTCATCTTGCAGGAGGATGATGATGATGAT 1620
Db 1561 TCGAAGAAACTGCCCGCAGAGATCAGAGTTCATCTTGCAGGAGGATGATGATGATGAT 1620
Qy 1621 CACGAAACAGTCCAAAGAAATGTGTCAGGCTGCGAGGAGGAGTGCAGTGCAGTTCGTT 1680
Db 1621 CACGAAACAGTCCAAAGAAATGTGTCAGGCTGCGAGGAGGAGTGCAGTGCAGTTCGTT 1680
Qy 1681 CCACTCTTACGAGACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTG 1740
Db 1681 CCACTCTTACGAGACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTG 1740
Qy 1741 TTATGA 1746
Db 1741 TTATGA 1746


```
QY 512 TGATCCTTGTAGATGACTACAGTGATAGAGAGACCTGGAAGAGCGCTTGCCCAATGAGC 571
Db 307 TGATCCTTGTAGATGACTACAGTGATAGAGAGACCTGGAAGAGCGCTTGCCCAATGAGC 366
QY 572 TTTGGGAGCTGCCAAGGTGCGCTGATGCTCCGCCCAACAAGAGAGAGCGCTGGTGGAG 631
Db 367 TTTGGGAGCTGCCAAGGTGCGCTGATGCTCCGCCCAACAAGAGAGAGCGCTGGTGGAG 426
QY 632 CCCGGCTGCTGGGGCGCTCGCGGAGGGCGGATGTTCTGACCTTCTGGAGCTGTCACT 691
Db 427 CCCGGCTGCTGGGGCGCTCGCGGAGGGCGGATGTTCTGACCTTCTGGAGCTGTCACT 486
QY 692 GTGAGTCCAGAGAGGGTGGTGGAGCGCTGCTGCAGAGGATCCATGAAGAGAGTGG 751
Db 487 GTGAGTCCAGAGAGGGT-GCTGGAGCGCTGCTGCAGAGGATCCATGAAGAGAGTGG 545
QY 752 CAGTGGTGTGCCCGGTGATTGATGTGATCGACTGGAACACCTTCGAATACCTGGGAACT 811
Db 546 CAGTGGTGTGCCCGGTGATTGATGTGATCGACTGGAACACCTTCGAATACCTGGGAACT 605
QY 812 CCGGGAGCCCCAGATCGCGGTTTTCGACTGGAGGCTGGTGTTCACGTGGCACACAGTTC 871
Db 606 CCGGGAGCCCCAGATCGCGGTTTTCGACTGGAGGCTGGTGTTCACGTGGCACACAGTTC 665
QY 872 CTGAGAGGAGAGATACGATGCAATCCCGCTGATGTCATCAGGTCTCCAAATG 931
Db 666 CTGAGAGGAGAGATACGATGCAATCCCGCTGATGTCATCAGGTCTCCAAATG 725
QY 932 CTGTTGGGCTGTTTGGTGTGAGTGAAGAAATTTTGAATATCTGGGCTTATCATACAG 991
Db 726 CTGTTGGGCTGTTTGGTGTGAGTGAAGAAATTTTGAATATCTGGGCTTATCATACAG 785
QY 992 GAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTG 1051
Db 786 GAATGGAAGTTTGGGAGGAGAAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTG 845
QY 1052 GGGTCTTGGAACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCTCCCT 1111
Db 846 GGGTCTTGGAACACACCCATGTTCCCATGTTGGCCATGTTTCCCAAGCAAGCTCCCT 905
QY 1112 ACTCCGCAACAGGCTCTGCGCAACAGTGTCTGCGAGCTGAAGTATGGATGATGAAT 1171
Db 906 ACTCCGCAACAGGCTCTGCGCAACAGTGTCTGCGAGCTGAAGTATGGATGATGAAT 965
QY 1172 TTAAGAGCTCTACTACCATCGCAACCCCGTGGCCGCTTGGAACTTTTGGGATGTA 1231
Db 966 TTAAGAGCTCTACTACCATCGCAACCCCGTGGCCGCTTGGAACTTTTGGGATGTA 1025
QY 1232 CAGAGAGAGAGCTCGGGACAGCTCCAGTGTAAAGACTTCAAGTGTCTTGAGGA 1291
Db 1026 CAGAGAGAGAGCTCGGGACAGCTCCAGTGTAAAGACTTCAAGTGTCTTGAGGA 1085
QY 1292 CTGTGTATCCAGAACTGATGCTGCTGAGGACAGGCTTCTCGGGATGCTCCAGA 1351
Db 1086 CTGTGTATCCAGAACTGATGCTGCTGAGGACAGGCTTCTCGGGATGCTCCAGA 1145
QY 1352 ACAAGGACTAACAGACTACTGCTTTGACTATAACCTCCCGATGAAACACAGATTGTGG 1411
Db 1146 ACAAGGACTAACAGACTACTGCTTTGACTWTAACTCCCGATGAAACACAGATTGTGG 1205
QY 1412 GACACAGGCTATTCTGTACTCTGTCTGATGGATGGGCGAGAACTCA 1457
Db 1206 GACACAGGCTATTCTGTACTCTGTCTGATGGATGGGCGAGAACTCA 1251
```

RESULT 7

```
US-10-106-698-598
; Sequence 598, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
```

```
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 598
; LENGTH: 1517
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (144)..(144)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-598
```

```
Query Match 40.8%; Score 712; DB 14; Length 1517;
Best Local Similarity 100.0%; Pred. No. 1.6e-189;
Matches 712; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1035 GATCTGCGAGTGTGGTGGGTTCTTGGAAACACACCCATGTTCCCATGTTGCCCATGTTTT 1094
Db 479 GATCTGCGAGTGTGGTGGGTTCTTGGAAACACACCCATGTTCCCATGTTGCCCATGTTTT 538
QY 1095 CCCCAGCAGCTCCCTACTCTCCGCAACAAGGCTCTGCGCAACAGTGTTCGTCAGCTGA 1154
Db 539 CCCCAGCAGCTCCCTACTCTCCGCAACAAGGCTCTGCGCAACAGTGTTCGTCAGCTGA 598
QY 1155 AGTATGATGGATGAAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGGCCCTTGGGA 1214
Db 599 AGTATGATGGATGAAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGGCCCTTGGGA 658
QY 1215 ACCTTTTGGGATGTGACAGAGAGGAGAGCTCCGGGACAAAGCTCCAGTGTAAAGACTT 1274
Db 659 ACCTTTTGGGATGTGACAGAGAGGAGAGCTCCGGGACAAAGCTCCAGTGTAAAGACTT 718
QY 1275 CAAGTGGTCTTGGAGACTGTGATCCAGAACTGTCATGTCGTCGAGGACAGGCTTGGCTT 1334
Db 719 CAAGTGGTCTTGGAGACTGTGATCCAGAACTGTCATGTCGTCGAGGACAGGCTTGGCTT 778
QY 1335 CTTCCGGGATGCTCCAGAAACAAAGGACTTAAACAGACTACTGCTTTGACTATACCTCCCGA 1394
Db 779 CTTCCGGGATGCTCCAGAAACAAAGGACTTAAACAGACTACTGCTTTGACTATACCTCCCGA 838
QY 1395 TGAACACCAAGTGTGGGACACCAAGGCTCATTTCTGTACCTCTGTCTATGGGATGGGCCAGAA 1454
Db 839 TGAACACCAAGTGTGGGACACCAAGGCTCATTTCTGTACCTCTGTCTATGGGATGGGCCAGAA 898
QY 1455 TCAGTTTTTCGATACACGTCCAGAAAGAAATACGCTATTAACCCACAGGCTTGGGG 1514
Db 899 TCAGTTTTTCGATACACGTCCAGAAAGAAATACGCTATTAACCCACAGGCTTGGGG 958
QY 1515 CTGCATTGTGTGAAGCAGGAATGGAATCCCTTATCATCATCTCTGTCTGTCGGAAGAACTGC 1574
Db 959 CTGCATTGTGTGAAGCAGGAATGGAATCCCTTATCATCATCTCTGTCTGTCGGAAGAACTGC 1018
QY 1575 CCCAGAGAATCAGAAGTTCACTTTGCGAGGAGGATGGATCTTTTATTTTCCAGAAAGTCCAA 1634
Db 1019 CCCAGAGAATCAGAAGTTCACTTTGCGAGGAGGATGGATCTTTTATTTTCCAGAAAGTCCAA 1078
QY 1635 GAAATGTCTCAGGCTCGGAGGAGGAGTGCAGTGCAGTGTTCGTTCCCTCTTTACGAGA 1694
Db 1079 GAAATGTCTCAGGCTCGGAGGAGGAGTGCAGTGCAGTGTTCGTTCCCTCTTTACGAGA 1138
QY 1695 CTGCACCAACTCCGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGA 1746
Db 1139 CTGCACCAACTCCGATCATCAGAAATGTTCTTCAAGAGCGCATGTTATGA 1190
```


RESULT 8

US-09-895-298-44
; Sequence 44, Application US/09895298
; Publication No. US20030078405A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 47 Human Secreted Proteins
; FILE REFERENCE: P2035P1
; CURRENT APPLICATION NUMBER: US/09/895,298
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/591,16
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: PCT/US99/29950
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: 60/113,006
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 60/112,809
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 1517
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (144)
; OTHER INFORMATION: n equals a,t,g, or c

US-09-895-298-44

Query Match 40.8%; Score 711.6; DB 11; Length 1517;
Best Local Similarity 99.9%; Pred. No. 2.1e-189;
Matches 711; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1035	GATCTGGCAGTGTGGTGGGTTCTGGAAACACACCATGTTCCCATGTGGCCATGTTT	1094
Db	479	GATCTGGCAGTGTGGTGGGTTCTGGAAACACACCATGTTCCCATGTGGCCATGTTT	538
Qy	1095	CCCCAAGCAAGCTCCCTACTCCCGCAACAAGCTCTGCCCAACAGTGTTCGTGAGCTGA	1154
Db	539	CCCCAAGCAAGCTCCCTACTCCCGCAACAAGCTCTGCCCAACAGTGTTCGTGAGCTGA	598
Qy	1155	AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGGCCCTTGGGA	1214
Db	599	AGTATGGATGGATGAATTTAAAGAGCTCTACTACCATCGCAACCCCGTGGCCCTTGGGA	658
Qy	1215	ACCTTTTGGGATGTGACAGAGAGAAAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTT	1274
Db	659	ACCTTTTGGGATGTGACAGAGAGAAAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTT	718
Qy	1275	CAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCTTGCTT	1334
Db	719	CAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCTTGCTT	778
Qy	1335	CTTCGGGATGCTCCAGAAACAAAGGACTAAACAGACTTCTGTGACTATAACCCCTCCCGA	1394
Db	779	CTTCGGGATGCTCCAGAAACAAAGGACTAAACAGACTTCTGTGACTATAACCCCTCCCGA	838
Qy	1395	TGAAACACAGATTTGGGACACCAAGGTCAATCTGTACCTCTGTCTATGGATGGGCCAGAA	1454
Db	839	TGAAACACAGATTTGGGACACCAAGGTCAATCTGTACCTCTGTCTATGGATGGGCCAGAA	898
Qy	1455	TCAGTTTTTCAGTACACGTCCTCAGAAAGAAATAGCTTATACACCCACAGCCTGAGGG	1514
Db	899	TCAGTTTTTCAGTACACGTCCTCAGAAAGAAATAGCTTATACACCCACAGCCTGAGGG	958
Qy	1515	CTGCATTCTGTGGAAAGCAGGAATGGATACCCCTTATCATGTCTCTCTCGAAGAACTGC	1574
Db	959	CTGCATTCTGTGGAAAGCAGGAATGGATACCCCTTATCATGTCTCTCTCGAAGAACTGC	1018
Qy	1575	CCCAGAGAAATCAGAAATTCATCTTTCAGAGAGATGGATCTTTTATTTACGAAACAGTCCAA	1634
Db	1019	CCCAGAGAAATCAGAAATTCATCTTTCAGAGAGATGGATCTTTTATTTACGAAACAGTCCAA	1078

Qy	1635	GAATGTGTCCAGGCTCGAGGAAGAGTCCAGTGCAGTTCGTTCCACTCTTACGAGA	1694
Db	1079	GAATGTGTCCAGGCTCGAGGAAGAGTCCAGTGCAGTTCGTTCCACTCTTACGAGA	1138
Qy	1695	CTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA	1746
Db	1139	CTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA	1190

RESULT 9

US-10-292-896-121
; Sequence 121, Application US/10292896
; Publication No. US20030186850A1
; GENERAL INFORMATION:
; APPLICANT: HASSAN, Helle
; APPLICANT: REIS, Celso A.
; APPLICANT: BENNETT, Eric P.
; APPLICANT: CLAUSEN, Henrik
; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GA
; TITLE OF INVENTION: TRANSPERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS
; TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICAMENTS
; FILE REFERENCE: 4305/1H154-US3
; CURRENT APPLICATION NUMBER: US/10/292,896
; PRIOR FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/425,204
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/DK01/00328
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 60/203,331
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 121
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-292-896-121

Query Match 30.0%; Score 524; DB 12; Length 525;
Best Local Similarity 100.0%; Pred. No. 7.3e-137;
Matches 524; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1223	GGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGT	1282
Db	2	GGGATGTGACAGAGAGGAAGCAGCTCCGGGACAAAGCTCCAGTGTAAAGACTTCAAGTGT	61
Qy	1283	TCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCTTGCTTTCGGGA	1342
Db	62	TCTTGGAGACTGTGTATCCAGAACTGCATGTGCTGAGGACAGGCTTGCTTTCGGGA	121
Qy	1343	TGCTCCAGAAACAAAGGACTAACAGACTACTGCTTTGACTATTAACCTCCCGATCAAAACC	1402
Db	122	TGCTCCAGAAACAAAGGACTAACAGACTACTGCTTTGACTATTAACCTCCCGATCAAAACC	181
Qy	1403	AGATTGTGGACACCAAGTCAATCTGTACCTCTGTATGGATGGGCGAGAACTCAGTTTT	1462
Db	182	AGATTGTGGACACCAAGTCAATCTGTACCTCTGTATGGATGGGCGAGAACTCAGTTTT	241
Qy	1463	TCGAGTACAGTCCCGAAGAAATACGCTATAACCCACAGCCTGAGGGCTGCATTG	1522
Db	242	TCGAGTACAGTCCCGAAGAAATACGCTATAACCCACAGCCTGAGGGCTGCATTG	301
Qy	1523	CTGTGGAGCAGGATGATACCCCTTATCATGCTCTCTCGGAGAAACTGCCCCAGAGA	1582
Db	302	CTGTGGAGCAGGATGATGATACCCCTTATCATGCTCTCTCGGAGAAACTGCCCCAGAGA	361
Qy	1583	ATCAGAAATTCATCTTCAGAGGATGATCTTTTATTTTACGAAACAGTCCAAAGAAATGTG	1642
Db	362	ATCAGAAATTCATCTTCAGAGGATGATCTTTTATTTTACGAAACAGTCCAAAGAAATGTG	421
Qy	1643	TCCAGGCTGCGAGGAAGAGTCCAGTGCAGATTCGTTTCCATCTTACGAGACTGCACCA	1702

Db 422 TCCAGCTCGCAGGAGGAGTGCAGTGACAGTTTCGTTCCACTCTTACGAGACTGCACCA 481
Qy 1703 ACTCGATCATCAAGAAATGGTCTTCAAGAGCGCATGTTATGA 1746
Db 482 ACTCGATCATCAAGAAATGGTCTTCAAGAGCGCATGTTATGA 525

RESULT 10

US-09-867-701-4790
; Sequence 4790, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4790
; LENGTH: 496
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-4790

Query Match 25.8%; Score 450.8; DB 10; Length 496;
Best Local Similarity 96.8%; Pred. No. 2.7e-116; Indels 4; Gaps 2;
Matches 482; Conservative 0; Mismatches 12;
Qy 203 CGCGGTGCCGCGAAGCGCTGGCGCGCGGCGGCGAGCGGTGCGGTGCGAGTGAATCACTACCTC 262
Db 2 CGCGCGGTGCCGCGAAGCGCTGGCGG--GGAGCGAGGCGGTGCGGTGCGAGTGAATCACTACCTC 58
Qy 263 GCGAGGAGCTCGCGTGCAGAG--GAGAGCTGGCGTGCACCGAGTGAATCACTACCTC 321
Db 59 GCGAGGAGCTCGCGTGCAGAGAGAGAGCGTGCAGTGCACCGAGTGAATCACTACCTC 118
Qy 322 AGCGACCGCATCTCACTGACCGCGCTGCCGCGCGTGCAGCGCTGGAACCGCTGTGCAAGAG 381
Db 119 AGCGACCGCATCTCACTGACCGCGCTGCCGCGCGTGCAGCGCTGGAACCGCTGTGCAAGAG 178
Qy 382 AAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAA 441
Db 179 AAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTTATAATGAA 238
Qy 442 GCCTGGTCACTCTCTTCGGACAGTTTACAGTGTCTTGAGACATCCCGGATATCTCTG 501
Db 239 GCCTGGTCACTCTCTTCGGACAGTTTACAGTGTCTTGAGACATCCCGGATATCTCTG 298
Qy 502 CTAGAAGAGTATCTCTTGTAGATGACTACAGTGTATAGAGACCTGAGAGCGCTTG 561
Db 299 CTAGAAGAGTATCTCTTGTAGATGACTACAGTGTATAGAGACCTGAGAGCGCTTG 358
Qy 562 GCCAATGAGCTTTCGGACCTGCCAAGGTGCGCTGATCCCGCGCAACAAGAGAGGGC 621
Db 359 GCCAATGAGCTTTCGGACCTGCCAAGGTGCGCTGATCCCGCGCAACAAGAGAGGGC 418
Qy 622 CTGTGGAGCGCGGTGCTGGGGCGTCTCGCGGAGGGCGATGTTCTGACCTTCTG 681
Db 419 CTGTGGAGCGCGGTGCTGGGGCGTCTCGCGGAGGGCGATGTTCTGACCTTCTG 478
Qy 682 GACTGTCACTGTAGTGC 699
Db 479 GACTGTCACTGTAGTGC 496

RESULT 11

US-09-867-701-4789/c
; Sequence 4789, Application US/09867701
; Patent No. US20020132237A1

; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4789
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-4789

Query Match 25.4%; Score 443.4; DB 10; Length 473;
Best Local Similarity 99.4%; Pred. No. 3.1e-114;
Matches 466; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
Qy 595 CTGATCCGCGCAACAAGAGAGAGGCGCTGGTGGAGCCCGGCTGCTGGGGGCTCTGCG 654
Db 473 CTTATCCGCGCAACAAGAGAGAGGCGCTGGTGGAGCCCGGCTGCTGGGGGCTCTGCG 414
Qy 655 GCGAGGGCGGATGTTCTGACCTTCTTGACTGTCTGAGTGCACGAGGGTGGCTG 714
Db 413 GCGAGGGCGGATGTTCTGACCTTCTTGACTGTCTGAGTGCACGAGGGT-GCTG 355
Qy 715 GAGCCGCTGCTGCAGAGGATCCATGAAGAGAGTCCGAGTGGTGGCCGCTGATTGAT 774
Db 354 GAGCCGCTGCTGCAGAGGATCCATGAAGAGAGTCCGAGTGGTGGCCGCTGATTGAT 295
Qy 775 GTGATCGACTGGAAACACCTTCGAAATACCTCGGGAACCTCCGGGAGCCCGAGTGGCGGT 834
Db 294 GTGATCGACTGGAAACACCTTCGAAATACCTCGGGAACCTCCGGGAGCCCGAGTGGCGGT 235
Qy 835 TTCGACTGGAGGCTGGTGTTCAGC-TGGCACACAGTTTCTTGAGAGGAGAGGATACCGAT 893
Db 234 TTCGACTGGAGGCTGGTGTTCAGCCTGCGCACACAGTTTCTTGAGAGGAGAGGATACCGAT 175
Qy 894 GCAATCCCGCTGCAGTGTCTCAGGTCTCCAAATGGCTGGTGGGCTGTTTGTGTGAG 953
Db 174 GCAATCCCGCTGCAGTGTCTCAGGTCTCCAAATGGCTGGTGGGCTGTTTGTGTGAG 115
Qy 954 TAAGAAATATTTGAATATCTGGGCTCTTATGATACAGGAATGGAAGTTTGGGAGGAGA 1013
Db 114 TAAGAAATATTTGAATATCTGGGCTCTTATGATACAGGAATGGAAGTTTGGGAGGAGA 55
Qy 1014 AAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAA 1062
Db 54 AAACCTCGAATTTTCTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAA 6

RESULT 12

US-09-867-701-3683/c
; Sequence 3683, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3683
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Homo sapien

US-09-867-701-3683

Query Match 24.4%; Score 425.8; DB 10; Length 473;
Best Local Similarity 98.7%; Pred. No. 2.8e-109;
Matches 471; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

Qy 592 CGCTGATCCGCGCAACAAGAGAGAGGGCTGGTGGAGCCGGCTGCTGGGGCGTCT 651
Db 473 CGCTGATCCGCGCAACAAGAGAGAGGGCTGGTGGAGCCGGCTGCTGGGGCGTCT 414

Qy 652 CGCGGAGGGCGGATGTTCTGACCTTCTGACCTGTCTGACCTGTCTGAGTGCACGAAGGGTGG 711
Db 413 CGCGGGA-GGGCGATGTTCTGACCTTCTGACCTTCTGACCTGTCTGAGTGCACGAAGGGT-G 356

Qy 712 CTGAGCGCGCTGCTGAGAGGATCCATGAAGAGAGTCCGAGAGTGTGTCGCCGTGATT 771
Db 355 CTGAGCGCGCTGCTGAGAGGATCCATGAAGAGAGTCCGAGAGTGTGTCGCCGTGATT 296

Qy 772 GATGTGATCGACTGGAAACACCTTCCGAATACCTGGGGAATCCCGGGAGCCCCAGATCGGC 831
Db 295 GATGTGATCGACTGGAAACACCTTCCGAATACCTGGGGAATCC-CGGAGCCCCAGATCGGC 237

Qy 832 GGTTCGATCGAGGCTGGTGTTCAGTGGCACACAGTTCCTGAGAGGAGGATACGG 891
Db 236 GGTTCGATCGAGGCTGGTGTTCAGTGGCACACAGTTCCTGAGAGGAGGATACGG 177

Qy 892 ATGCAATCCCGCTGATGTCATCAGGTCTCCAAATGCTGCTGGGCTGTTGCTGTG 951
Db 176 ATGCAATCCCGCTGATGTCATCAGGTCTCCAAATGCTGCTGGGCTGTTGCTGTG 117

Qy 952 AGTAAGAAATATTTGAAATATCTGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGA 1011
Db 116 AGTAAGAAATATTTGAAATATCT-GGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGA 58

Qy 1012 GAAACCTCGAATTTCTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAACACAC 1068
Db 57 GAAACCTCGAATTTCTTTAGGATCTGGCAGTGTGGTGGGTTCTGGAACACAC 1

RESULT 13

US-09-777-564-414
; Sequence 414, Application US/09777564
; Patent No. US2002022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 414
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(544)
; OTHER INFORMATION: n = A,T,C or G

US-09-777-564-414
Query Match 24.0%; Score 418.4; DB 9; Length 544;
Best Local Similarity 99.8%; Pred. No. 3.5e-107;
Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1327 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGACTTAACAGACTACTGCTTTGACTATAAC 1386
Db 1 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGACTTAACAGACTACTGCTTTGACTATAAC 60

Qy 1387 CCTCCGATGAAACCAAGATTGGGACACCAAGTCACTTCTGTACCTCTGTATGGGATG 1446

Db 61 CTCTCCGATGAAACCAAGATTGGGACACCAAGTCACTTCTGTACCTCTGTATGGGATG 120
Qy 1447 GGCAGAAATCAGTCTTTTCGAGTACAGCTCCAGAGAAATAGCTATATAACACCCACAG 1506
Db 121 GGCAGAAATCAGTCTTTTCGAGTACAGCTCCAGAGAAATAGCTATATAACACCCACAG 180
Qy 1507 CTTGAGGGCTGCAATGCTGTGGAGCAGGAATCGATACCTTATCATGCACTCTCTGCGAA 1566
Db 181 CTTGAGGGCTGCAATGCTGTGGAGCAGGAATCGATACCTTATCATGCACTCTCTGCGAA 240

Qy 1567 GAAACTGCCCCAGAGAAATCAGAAATTCATCTTCAGAGAGATGGATCTTTTATTTACGAA 1626
Db 241 GAAACTGCCCCAGAGAAATCAGAAATTCATCTTCAGAGAGATGGATCTTTTATTTACGAA 300

Qy 1627 CAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAAGAGTCCAGTCTCGTTCCACTC 1686
Db 301 CAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAAGAGTCCAGTCTCGTTCCACTC 360

Qy 1687 TTACGAGACTGCACCAACTCCGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 1746
Db 361 TTACGAGACTGCACCAACTCCGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 420

RESULT 14

US-10-015-219-414
; Sequence 414, Application US/10015219
; Publication No. US20030008299A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493C1
; CURRENT APPLICATION NUMBER: US/10/015,219
; CURRENT FILING DATE: 2002-03-02
; NUMBER OF SEQ ID NOS: 1739
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 414
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 544
; OTHER INFORMATION: n = A,T,C or G

US-10-015-219-414
Query Match 24.0%; Score 418.4; DB 14; Length 544;
Best Local Similarity 99.8%; Pred. No. 3.5e-107;
Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1327 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGACTTAACAGACTACTGCTTTGACTATAAC 1386
Db 1 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGACTTAACAGACTACTGCTTTGACTATAAC 60

Qy 1387 CCTCCGATGAAACCAAGATTGGGACACCAAGTCACTTCTGTACCTCTGTATGGGATG 1446
Db 61 CCTCCGATGAAACCAAGATTGGGACACCAAGTCACTTCTGTATGGGATG 120

Qy 1447 GGCAGAAATCAGTCTTTTCGAGTACAGCTCCAGAGAAATAGCTATATAACACCCACAG 1506
Db 121 GGCAGAAATCAGTCTTTTCGAGTACAGCTCCAGAGAAATAGCTATATAACACCCACAG 180

Qy 1507 CTTGAGGGCTGCAATGCTGTGGAGCAGGAATGGATACCTTATCATGCACTCTCTGCGAA 1566
Db 181 CTTGAGGGCTGCAATGCTGTGGAGCAGGAATGGATACCTTATCATGCACTCTCTGCGAA 240

Qy 1567 GAAACTGCCCCAGAGAAATCAGAAATTCATCTTCAGAGAGATGGATCTTTTATTTACGAA 1626
Db 241 GAAACTGCCCCAGAGAAATCAGAAATTCATCTTCAGAGAGATGGATCTTTTATTTACGAA 300

Qy 1627 CAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAAGAGTCCAGTCTCGTTCCACTC 1686
Db 301 CAGTCCAAAGAAATGTGTCCAGGCTGCGAGGAAGAGTCCAGTCTCGTTCCACTC 360

Qy 1687 TTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 1746
Db 361 TTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATGA 420

RESULT 15

US-09-777-564-172
; Sequence 172, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
; APPLICANT: Mannion, Paul A.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 172
; LENGTH: 632
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(632)
; OTHER INFORMATION: n = A,T,C or G
US-09-777-564-172

Query Match 22.9%; Score 400.2; DB 9; Length 632;
Best Local Similarity 98.1%; Pred. No. 4.9e-102;
Matches 413; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
Qy 1327 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTTAACAGACTACTGCTTTTGACTATAAC 1386
Db 1 CCTGGCTTCTTCGGGATGCTCCAGAACAAAGGACTTAACAGACTACTGCTTTTGACTATAAC 60
Qy 1387 CCTCCCGATGAACACCAAGATTGTGGGACACAGGTCTATCTGTATCTCTGTATGGGATG 1446
Db 61 CCTCCCGATGAACACCAAGATTGTGGGACACAGGNCATCTGTACCTCTGTATGGGATG 120
Qy 1447 GGCCAGATC-AGTTTTTCGAGTACAGTCCAGAAAGAAATACGCTATAACCCACCA 1505
Db 121 GGCCAGATCAAGTTTTTCGAGTACACTTCCAGAAAGAAATACGCTATAACCCACCA 180
Qy 1506 GCCTGAGGGCTGCTATGCTGTGGAAGCAGGAATGATACCTTATCATGCTCTCTCGGA 1565
Db 181 GCCTGAGGGCTGCTATGCTGTGGAAGCAGGAATGATACCTTATCATGCTCTCTCGGA 240
Qy 1566 AGAACTGCCCCAGAGAAATCAGAAAGTTTCATCTTCAGGAGGATGGATCTTTTATTTACGA 1625
Db 241 AGAACTGCCCCAGAGAAATCAGAAAGTTTCATCTTCAGGAGGATGGATCTTTTATTTACGA 300
Qy 1626 ACAGTCAAGAAATGCTGTCAGGCTGGAGGAAGGATCGAGTGACAGTTTCGTTCCACT 1685
Db 301 ACAGTCAAGAAATGCTGTCAGGCTGGAGGAAGGATCGAGTGACAGTTTCGTTCCACT 360
Qy 1686 CTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATG 1745
Db 361 CTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATGTTATG 420
Qy 1746 A 1746
Db 421 A 421

Search completed: December 8, 2003, 01:56:05
Job time : 534.474 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 12:46:13 ; Search time 3629.14 Seconds
(without alignments)
11693.015 Million cell updates/sec

Title: US-10-074-527-3

Perfect score: 1746

Sequence: 1 atgtggggggcgacggcg.....tcaaagagcgcatgttatga 1746

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

BST.*

1: em_estba.*
2: em_estum.*
3: em_estin.*
4: em_estm.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_man.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1249.4	71.6	2074	11 AK042133	AK042133 Mus muscu
2	1249.4	71.6	2217	11 AK033638	AK033638 Mus muscu
3	1238.4	70.9	2034	11 AK082014	AK082014 Mus muscu
4	1025.4	58.7	1781	11 BC025639	BC025639 Mus muscu

5	936.2	53.6	2064	11 BC024475	BC024475 Mus muscu
6	925.2	53.0	1752	11 AK043977	AK043977 Mus muscu
7	660.8	37.8	676	10 BE677813	BE677813 7f63d08.x
c	566.4	32.4	692	14 CB165548	CB165548 BQ603160
8	548.2	31.4	561	9 AI800923	AI800923 wgl4f02.x
10	520.8	29.8	2607	11 AK033494	AK033494 Mus muscu
11	519	29.7	967	10 BG167520	BG167520 602342726
12	508	29.1	661	12 BI154821	BI154821 602902801
13	503.2	28.5	721	12 BG868185	BG868185 602787291
14	480.6	27.8	511	9 AI863865	AI863865 wj5b07.x
15	450.8	25.8	496	9 AA429394	AA429394 zw32h12.g
16	448.2	25.7	700	13 BU307661	BU307661 603405492
c	443.4	25.4	473	9 AA429393	AA429393 zw32h12.x
18	426.2	24.4	530	10 BG608482	BG608482 307084 MA
c	425.8	24.4	473	9 AA401053	AA401053 zu50f12.x
20	424.8	24.3	653	10 BG625226	BG625226 BB625226
c	424.2	24.3	457	12 BG993195	BG993195 RC4-HT109
21	413	23.7	664	10 BB638950	BB638950 BB638950
22	387.4	22.2	477	10 BG608487	BG608487 307090 MA
23	385.8	22.1	628	10 BB649873	BB649873 BB649873
24	373.4	21.4	521	10 BE647828	BE647828 UI-M-BH1-
25	373	21.4	544	4 BX514155	BX514155 RZPD Mus
26	365.4	20.9	697	10 BB272149	BB272149 BB272149
27	352.8	20.2	772	10 BF247687	BF247687 601858772
28	347.4	19.9	703	12 BI663830	BI663830 603288578
30	346.8	19.9	462	14 CD554038	CD554038 BQ370807-
31	340	19.5	445	10 BF211275	BF211275 601812512
32	335	19.2	407	12 BI403200	BI403200 MT-P-CP1-
33	333.8	19.1	909	14 CD251293	CD251293 AGENCOURT
34	332.2	19.0	441	9 AW462202	AW462202 BP230008B
35	331.8	19.0	464	13 BQ829759	BQ829759 LL61n2066
36	329.6	18.9	458	10 BE138172	BE138172 ug49d03.y
37	325	18.6	445	14 CB552134	CB552134 WMSPO054
38	323	18.5	690	10 BF031122	BF031122 601558805
39	322.2	18.5	360	9 AA493187	AA493187 ne56f02.s
40	320.4	18.4	442	12 BM254405	BM254405 S15827 MA
41	318.4	18.2	818	10 BF246502	BF246502 601854747
42	311.8	17.9	634	12 BJ067676	BJ067676 BJ067676
43	306.4	17.5	471	9 AA871181	AA871181 vq32b03.x
44	298	17.1	947	10 BG294191	BG294191 602391264
45	296	17.0	770	14 CD299534	CD299534 AGENCOURT

ALIGNMENTS

RESULT 1

AK042133

LOCUS

DEFINITION

AK042133

Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630062B03 product:hypothetical Glycosyl transferase, family 2 containing protein, full insert sequence.

AK042133

GI:26334956

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1

Carninci, P., Shibata, Y., Hayata, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

9279253

10349636

2

Carninci, P., Shibata, Y., Hayata, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

11042159

Db 519 GAGACCTCCCTGACATCTCTGCTGAGAGAGGTCAATCTTGTTAGATGATCACTACAGGACAGA 578
Qy 541 GAGCACCTGAAGGAGCGCTTGGCCAAATGAGCTTTTCGGGACTGCCCAAGAGGTGCGCTGATC 600
Db 579 GAGCACCTAAAGGAACGCTTGGCCAAACGAGCTGTACAGACTCCCAAGGTGGCGCTGATC 638
Qy 601 CGCGCCAAAGAGAGAGGCGCTGCTGCGAGACCGCGCTGCTGCGGCGCTGCTGCGGCGAG 660
Db 639 CGTGTACGAGGAGAGAGGCGCTAGTGTGAGACCGCGCTACTGCGGAGCGCTCTGCGGCGCAG 698
Qy 661 GCGGATGCTTCTGACCTTCTGAGCTGTCACTGTGAGTGCCACGAAGGCTGCTGAGCGCG 720
Db 699 GCGGAGTGTGACCTTCTGAGCTGTCACTGTGAATGTATGAGGGGTGCTGAGCGCT 758
Qy 721 CTGCTGAGAGGATCCATGAAGAGGAGTCCGCGAGTGTGCTGCGGCTGATTTGATGTGATC 780
Db 759 CTGCTGAGAGGATCCACGAGAGGAGTCCGCGAGTGTGCTGCGGCTGATTTGATGTGATC 818
Qy 781 GACTGGAACACCTTTCGAATACCTGCGGAACTCCCGGAGAGCCCAAGATCGCGGCTTTCGAC 840
Db 819 GACTGGAACACCTTTCGATACCTTGGCAACTCCGGAGAGCGCGAGATTTGGCGGCTTTGAC 878
Qy 841 TGGAGGCTGTGTTTCACTGTGCGACACAGTTTCTGAGAGGAGAGGATACGATGCAATCC 900
Db 879 TGGCGGCTGTGTTTCACTGTGCGACAGTGTGTTTCCCGAGCGGAGCGCGAGTGTGCGGTG 938
Qy 901 CCGTGTGATGTATCAGGTCTTCAACAAATGGCTGGGCTGTTTGTGTGAGTAAGAAA 960
Db 939 CCGATCGATGTATCAGGTCTTCAACAAATGGCTGGGCTGTTTGTGTGAGTAAGAGA 998
Qy 961 TATTTTGAATATCTGGGCTTATGATACAGGAATGAAGTTTGGGAGAGAGAAACCTC 1020
Db 999 TATTTTGAATATCTGGGCTTATGATACAGGAATGAAGTTTGGGAGAGAGAAACCTT 1058
Qy 1021 GAATTTCTTTAGGATCTGCGAGTGTGGGCTTCTGGAAACACACCACTGTTCCTCAT 1080
Db 1059 GAGTTCCTTTAGGATCTGCGAGTGTGGGCTTCTGGAAACACACCACTGTTCCTCAT 1118
Qy 1081 GTTGGCCATGTTTCCCAAGCAAGCTCCCTACTTCCCGCAACAGGCTCTGGCCAAACAGT 1140
Db 1119 GTGGCCACGCTTCCCTAGCAAGCTCCCTACTTCCCGTAGCAAGGCTTGGCCAAACAGT 1178
Qy 1141 GTTGTGCTGAGTGAATGATGATGAATTTAAAGAGCTCTACTACCACTGCAACCC 1200
Db 1179 GTCCGAGCTGAGAGTGTGATGATGAATTTAAAGAACTCTACTACCACTGCAACCC 1238
Qy 1201 CGTCCCGCTTGAACCTTTTGGGATGTACAGAGAGGAGCGAGCTCCGGCAAGCTC 1260
Db 1239 CAGGCGCGCTGGAACCTTTTGGGAGCTGACAGAGAGGAGAGCTTCCGGGCTTAAGCTC 1298
Qy 1261 CAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCAATGTGCTGAG 1320
Db 1299 CAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCAATGTGCTGAG 1358
Qy 1321 GACAGGCTGGCTTCTTGGGATGTCTCCAGAAAGGAGCTTAAAGAGCTACTGCTTTGAC 1380
Db 1359 GACAGGCTGGCTTCTTGGGATGTCTCCAGAAAGGAGCTTAAAGAGCTACTGCTTTGAC 1418
Qy 1381 TATAACCTCCGATGAAACAGATTTGTGGACACAGGCTCATTTGTGACCTCTGTCAT 1440
Db 1419 TAGAATCTCCCAATGAAACCAATTTGTGAAGGCGCACAGGCTCTTGTACCTCTGCGCAT 1478
Qy 1441 GCGATGGGCGAGAACTAGTTTTTTCAGTACAGCTCCAGAAAGAAATACGCTATAACACC 1500
Db 1479 GCGATGGGCTGAAACAGATTTTTCAGTATACGATACAGCTTAAAGAAATACGCTATAACACC 1538
Qy 1501 CACAGCTGAGGCTGATTTGCTGTGGAAGAGAGAAATGGAATACCTTTATCATGATCTC 1560
Db 1539 CGCCAGGCGGAGGCTGATCATCAGCTGAGGAGGAGGATACCTTTGTGATGATCTC 1598
Qy 1561 TGGAGAAATCTGCCCCAGAGAAATCAGAAATTCATTTGAGGAGGAGTGAATCTTTATTT 1620

Db 1599 TGCGGAGAGACCGTCCAGAGAACCCAGGAGTTTCATCTCAGAGGAGTGCACATTTAGTT 1658
Qy 1621 CACGAACAGTCCAGAAATGTGTCCAGGCTCGCAGGAAGAGTGCAGTGCAGTTTCGTT 1680
Db 1659 CACAAGCACAGCAGGAAATGTGTGGAGGCCACAGAGAGGTGTAGACAACGCGTTTGA 1718
Qy 1681 CCACTCTTAGGAGACTGCACCAACTCGGATCATCAGAAATGGTTCTTCAAAGAGCGCATG 1740
Db 1719 CCATCTTAGGAGACTGTACCAACTCAGATAACACAGGAGTGTCTTCAAAGAGCGCATG 1778
Qy 1741 T 1741
Db 1779 T 1779

AK033638 2217 bp mRNA linear HTC 05-DEC-2002
Mus musculus adult male cecum cDNA, RIKEN full-length enriched
library, clone:1910206E10 product:hypothetical Glycosyl
transferase, family 2 containing protein, full insert sequence.
AK033638
ACCESSION AK033638.1 GI:26329324
VERSION
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL 2049374
MEDLINE
PUBMED 11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Iehii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamanoto, R., Matsumoto, H., Sakeguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL 20530913
MEDLINE
PUBMED 11076861
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Iehii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuoka, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Staehli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boiffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L.,

Qy	961	TATTTTGAATCTGGGGTCTTATGATACAGGAATGGAAGTTTGGGGAGGAGAAAAACCTC	1020
Ds	967	TATTTTGATTATCTGGGGTCTTACGATACAGGAATGGAAGTCTGGGGAGGAGAAAAACCTT	1026
Qy	1021	GAATTTTCTTTAGGACTGGCAGTGTGGGGTCTTGGAAACACACCCCATGTCCTCAT	1080
Ds	1027	GAGTTCTCTTTAGGACTGGCAGTGTGGGGTCTTGGAAACACCCCATGTCCTCAT	1086
Qy	1081	GTTGGCCATGTTTCCCAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAACAGT	1140
Ds	1087	GTGGCCACAGTCTTCCCTAAGCAAGCTCCCTACTCCCGCAACAAGGCTCTGGCCAAACAGT	1146
Qy	1141	GTTCTGTGAGCTGAAGTATGATGGATGAATTTAAAGAGCTCTACTACCATTCGCAACCCC	1200
Ds	1147	GTCCGAGCTGCAGAGTGTGGATGGATGAATTTAAAGAACTCTACTACCAACCGCAATCCC	1206
Qy	1201	CGTCCCGCTTGGAACTTTTGGGGATGTGACAGAGGAGCAGCTCCGGCAAGCTC	1260
Ds	1207	CAGCCCGCTTGGAACTTTTGGGGATGTGACAGAGGAGCAGCTTGGGGCTTAAGCTC	1266
Qy	1261	CAGTGAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCCTGAG	1320
Ds	1267	CAGTGAAGACTTCAAGTGGTCTTGTAGATACAGTGTACCAGAACTGCAGTGCACAG	1326
Qy	1321	GACAGGCTGCTTCTTGGGATGCTCCAGAAACAAAGACTTAACAGACTACTGCTTTGAC	1380
Ds	1327	GACAGGCTGCTTCTTGGGATGCTCCAGAAACAAAGACTTAACAGACTACTGCTTTGAC	1386
Qy	1381	TATAACCTCCGATGAACACAGATGTGGGACACAGGTCATCTGTACCTCTGTCTCAT	1440
Ds	1387	TACATCTCTCCCAATGAACACCATGTTGAAGGCCACACAGGCTCTTGTACCTCTGCTCAT	1446
Qy	1441	GGGATGGGCCAGAACTCAGTCTTTCAGTACACGCTCCAGAAAGAAATACGCTATAACACC	1500
Ds	1447	GGGATGGGTCAACACAGTCTTTCAGTATACGACCCGGAAGAAATACGCTATAACACC	1506
Qy	1501	CACAGGCTCAGGCTGATGCTGTGGAGACAGGAATGGAATACCTTATCATCATCTC	1560
Ds	1507	CGCCAGCCGGAAGCTGCATCACCCTGGAGACGGAAGGATACCTTGTTCATGGATCTC	1566
Qy	1561	TGCGAAGAACTGCCAGAGAACTCAGAGTTCATCTTCGAGGAGGATGGATCTTTATT	1620
Ds	1567	TGCGAGAGACCGTCCCAAGAAACAGAGTTCATCTTCACAGGAGGATGGACATTTAGT	1626
Qy	1621	CACGAACAGTCCAGAAATGTCTCAGGCTCGAGGAGGAGTCCAGTGACAGTTTCGTT	1680
Ds	1627	CACAGCACAGGAAATGTCTGGAGCCACAGAGAGGTGTAGACACGGCTTTGCA	1686
Qy	1681	CCACTCTTACGAGACTGCACCACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATG	1740
Ds	1687	CCATCTTACGGGACTGTACCACTCAGATAACAGAGGTGTTCTTCAAGGAGCGCATG	1746
Qy	1741	T 1741	
Ds	1747	T 1747	
RESULT 3	AK082014	LOCUS	
DEFINITION	AK082014	2034 bp mRNA linear	HTC 05-DEC-2002
ACCESSION	AK082014	Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:C130094P07 product:hypothetical Glycoyl transferase, family 2 containing protein, full insert sequence.	
VERSION	AK082014.1	GI:26100345	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota: Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL

High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitesunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staib, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2034)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Ohashi, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

QY 1680 TCACACTCTTAGCGACTGCACCACTCGGATCATCAGAAATGGTTCTTCAAGAGCGCAT 1739
 Db 1731 ACCATCTTAGCGGACTGTACCAACTCAGATAACCAAGAGGTGGTTCTTCAAGAGCGCAT 1790
 QY 1740 GT 1741
 Db 1791 GT 1792

RESULT 4
 BC025639
 LOCUS
 DEFINITION Mus musculus, clone IMAGE:5322763, mRNA.
 ACCESSION BC025639
 VERSION BC025639.1 GI:22213056
 KEYWORDS HTC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1781)
 Direct Submission
 Submitted (06-MAR-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nhgri.nih.gov
 Akhtar, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Guin, X., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaapi, R.,
 Maduro, Q.L., Masiello, C., Maaberi, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Taurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 56 Row: j Column: 12
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 12963494
 This clone has the following problem: retained intron.

FEATURES
 source
 1. 1781
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5322763"
 /tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
 ductal carcinoma, 5 month old virgin mouse."
 /clone_lib="NCI CGAP Mam6"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 BASE COUNT 436 a 443 c 490 g 412 t
 ORIGIN

Query Match 58.7% Score 1025.4; DB 11; Length 1781;
 Best Local Similarity 83.2%; Pred. No. 1.5e-219;
 Matches 1167; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

QY 339 GCACCGCGCGCTCCCGAGCGCTGGAAACCCGCTGTGCAAGAGAAGAAATATGATTATGA 398

Db 66 GCCCGCACCCCTCCCTCAGTCTCCAAACCATGGGTGCAGAGAGGTGAAATACGATTATGA 125
 QY 399 TAATTTCGCCAGACATCTGTTTATCATAGCATTTTATTAAGAACCTTGGTCAACTCTCCT 458
 Db 126 TAACCTGCCCAAGACTTCTGTTGTCATAGCATTTTATTAAGAACCTTGGTCCACACTCT 185
 QY 459 TCGGACAGTTTACAGTGTCTTGGAGACATCCCGGATATCCTGCTAGAAAGATGATCTCT 518
 Db 186 TCGGACAGTTTACAGTGTCTTGGAGACCTCCCTGACATCCTGCTGGAGAGGTCACTCT 245
 QY 519 TGTAGATGACTACAGTATAGAGAGACACTGAGGAGCGCTTGGCCATATGAGCTTTCCGG 578
 Db 246 GGTAGATGACTACAGGACAGAGAGACCTTAAAGGAAACGCTTGGCCAAACGAGCTGTCA 305
 QY 579 ACTGCCCAAGGTGGCTGATCCGCGCAACAGAGAGAGGCGCTGCTGGAGAGCCCGCT 638
 Db 306 GCTCCCAAGGTGGCTGATCCGCTGCTAGCAGAGAGAGGCGCTAGTGGAGAGCCCGCT 365
 QY 639 GCTGGGGCGTCTGCGGCGAGGGCGATGTTCTGACCTTCTCGACTGTCACTGTGAGTG 698
 Db 366 ACTGGGAGGCTCTGCGGCGAGGGCGAAGTGTGCTGACCTTCTAGACTGTCACTGTGAATG 425
 QY 699 CCACGAAGGTGGCTCGAGCGCGCTGCGAGAGATCCATGAGAGAGAGTCCGCGAGTGGT 758
 Db 426 TCATGAGGGTGGCTGGAGCGCTCTGCTGACAGAGGATCCACGAGAGAGTCCGCGAGTCT 485
 QY 759 GTCCCGGTGATTGATGTGATCGACTGGAAACACTTTCGAATACCTGGGAACTCCCGGGA 818
 Db 486 GTGCCCTGTTATCGATGATGATGACTGGAAACACTTGTAGTACCTGGGCAACTCGGAGA 545
 QY 819 GCCCAGATCGCGGTTTTCAGCTGGAGGCTGTTTTCAGTGGCAGACAGTCTTCTCGAGAG 878
 Db 546 GCGCAGATTGGCGCTTTCAGCTGGCGGCTGGTATTCAGTGGCAGCTGTTCCCGCAGCG 605
 QY 879 GGAGAGATACGATGCAATCCCGCTCGATGTCATCAGTCTCCAAACATATGGTGTGGTG 938
 Db 606 GGAGCGGATCGATCGATCGGTGCGCGATCGATGTTATCAGTCTCCAACTATGGCTGGGG 665
 QY 939 GCTGTTGCTGTGAGTAAGAAATATTTTGAATATCTGGGTCTTATGATACAGAGATGGA 998
 Db 666 CTGTTGCTGTGAGTAAGAGATATTTTGAATATCTGGGTCTTACGATACAGAGATGGA 725
 QY 999 AGTTTGGGAGGAGAAACCTCGAAATTTTCTTTAGGATCTGGCAGTGTGGTGGGTCTCT 1058
 Db 726 AGTCTGGGAGGAGAAACCTTGAGTCTCTTTAGGATCTGGCAGTGTGGTGGCTCTCT 785
 QY 1059 GGAAACACACCATGTTCCCATGTTGGCCATGTTTTCCTCCCAAGCAAGCTCCCTACTCCCG 1118
 Db 786 GGAAACGACCCGCTGCTCCCATGTTGGGCCACAGTCTTCCCTTAAGCAAGCTCCCTATTCCCG 845
 QY 1119 CAACAAGGCTCTGGCCCAACAGTGTTCGTCAGCTGAAGTATGATGATGATGATGATGATG 1178
 Db 846 TAGCAAGGCTCTGGCCCAACAGTGTTCGAGTGTGAGTGTGAGTGTGATGATGATGATG 905
 QY 1179 GCTCTACTACCATCGCAACCCCGCTGCGCGCTTGGAAACCTTTTGGGGATGTGACAGAGAG 1238
 Db 906 ACTCTACTACCATCGCAATCCCGCGCTGCGCGCTTGGAAACCTTTTGGGGACGTGACAGAGAG 965
 QY 1239 GAAGAGCTCTGGGGAACAAGCTCCAGTGTAAAGACTTCAAGTGTGTTCTTGGAGACTGTGTA 1298
 Db 966 GAAGAAGCTCTGGGCTTAAGCTCCAGTGTAAAGACTTCAAGTGTGTTCTTGTAGATACAGTGA 1025
 QY 1299 TCCAGAACTGTCATGCTGAGGAGAGGCTGCTTCTTGGGATGCTCCAGAACAAAGG 1358
 Db 1026 CCAGAAATGTCAGTGTCCAGAGAGAGGCTGCTTCTTGGGATGCTCCAGAACAAAGG 1085
 QY 1359 ACTTAACAGACTACTGCTTTGACTATTAACCTCCCGATGAAACAGAGTGTGGGACACCA 1418
 Db 1086 CTTAAGAGGCTACTGCTTGTGACTACATCTCCCATGAAACCATGTTGGAAGCCACCA 1145
 QY 1419 GGTCACTTGTACCTCTGTGATGGGATGGGCCAGAAATCAGTGTTCGAGTACACGTCCCA 1478

```

Db 1146 GGTCTTCTGTGATCTGTCCATGGGATGGGTGAGACCAACAGTCTTTTCGAGTATACGACCG 1205
QY 1479 GAAGAATAACGCTATTAACACCCACAGCCTGAGGGCTGCTGCTGCTGGAACAGGAAT 1538
Db 1206 GAAGAATAACGCTATTAACACCCACAGCCTGAGGGCTGCTGCTGCTGGAACAGGAAT 1265
QY 1539 GGATACCCCTTATCATGATCTCTGCGAAGAACTGCGCCAGAGAAATCAGAAAGTTTCATCTT 1598
Db 1266 GGATACCCCTTATCATGATCTCTGCGAAGAACTGCGCCAGAGAAATCAGAAAGTTTCATCTT 1325
QY 1599 GCAGGAGGATGATCTTTATTTTCAAGAACAGTCAAGAAATGTTCCAGCTCGAGGAA 1658
Db 1326 ACAGGAGGATGATCTTTATTTTCAAGAACAGTCAAGAAATGTTCCAGCTCGAGGAA 1385
QY 1659 GGATGAGTACAGTCTGTTGTTTCAAGAACAGTCAAGAAATGTTCCAGCTCGAGGAA 1718
Db 1386 GGTGTTAGAACAGGCTTTGACCATCTTACCGGACTGTACCAACTCAGATAACCCAGAG 1445
QY 1719 ATGTTTCTTCAAGAGCGCATGT 1741
Db 1446 GTGTTTCTTCAAGAGCGCATGT 1468

```

RESULT 5

```

BC024475 2064 bp mRNA linear HTC 04-MAR-2003
LOCUS Mus musculus, clone IMAGE:4976620, mRNA.
DEFINITION BC024475
ACCESSION BC024475.1 GI:22137472
VERSION HTC.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Strausberg, R.
Direct Submission
Submitted (01-MAR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

```

NIH-MGC Project URL: http://mgc.nci.nih.gov

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: http://www-shgc.stanford.edu

Contact: (Dickson, Mark) mcdpaxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAP Plate: 59 Row: J Column: 15
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis

This clone has the following problem: retained intron.

FEATURES

source

```

Location/Qualifiers
1..2064
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4976620"
/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI CGAP_Mam6"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"

```

```

BASE COUNT 475 a 505 c 625 g 459 t
ORIGIN
Query Match 53.6%; Score 936.2; DB 11; Length 2064;
Best Local Similarity 74.8%; Pred. No. 1.8e-199; Indels 185; Gaps 3;
Matches 1303; Conservative 0; Mismatches 253;
QY 1 ATGTGGGGCGCACGCGGGCGGCGCTGCGCCGCGGAACTGCGCGCGCGCGGAGGCG 60
Db 23 ATGTGGGGCGGCTGCGGTGAGGAGCGCTGCGCGCGGGGGCTGCGCGCGCGGAGGCG 82
QY 61 CTGTTGGTGTCTCTGCGCGCTGCTGCGGTGCGCGCGGCTGCGGTGCGGTGCGGCGCAG 120
Db 83 TTG-----CTGGCACTGCTGGCCCTGGCAGGGCTGGCGCGCTGCTGCTGCGGCG--- 129
QY 121 CGTGGGGCGGGGCGGGGCTGCGGAGCGGGGACCCCGCGCACCCCGCGCGCGCGGGCGG 180
Db 130 --TCGAAGCAGGTGCGGGACCGGTGGACCCCGGGACCCCGCGGACTCTCTCTCCCGGGCGT 187
QY 181 CGCGAGCGGTCTATGCGCGGGCGCGCTGCGCGGCGAAGCGCTGGGCGCGCGGGGCGAG 240
Db 188 CACGAGCCAGTGTGCGCCAGGCGCGCTGCGCGCGCGATGCTCTGGAGGCGCATGGGAG 247
QY 241 GCGGTGCGGTGCGAGCTGCGAGGCGGAGAGCTCGGCTGCGAGGAGAGAGCGTGCAGGCTG 300
Db 248 GCGGTGCGGTGCGAGCTGCGAGGCGGAGAGCTCGGCTGCGAGGAGAGAGCGTGAAGCAG 307
QY 301 CACCAGATTAACATCTTACCTCAGCAGCGGCGATCTCACTGACCGCGCTGCGCGAGCGC 360
Db 308 CACCAGATTAACATCTTATCTGAGCGACCGGTATCTGCTGCGCGCGGACTGCGCGAGCGC 367
QY 361 TGAACCCCGCTGTGCAAGAGAGAAATATGATTATGATAATTTTCCCGAGGACATCTGTT 420
Db 368 TGAACCCCTCT----- 378
QY 421 ATCATAGCATTTTATAATGAAGCCTGGTCAACTCTCTCTCGGACAGTTTACAGTGTCTT 480
Db 379 ----- 378
QY 481 GAGACATCCCGGATATCTCTGTAGAGAGTGTATCTTGTAGTACTACAGTGTAGATA 540
Db 379 ----- 378
QY 541 GAGCAGCTGAAGGAGGCGCTTGGGCAATAGCTTTTGGGAGCTGCCAAGGTGCGGCTGATC 600
Db 379 -AGCACTTAAGAGAACGCTTGGCCACAGAGCTGTACAGCTTCCCAAGGTGCGGCTGATC 437
QY 601 CGCGCCCAACAGAGAGAGGCGCTGCTGCGAGCGCGGCTGCTGGGGGCGCTGCGCGCGAGG 660
Db 438 CGTGTAGCAGGAGAGAGGCGCTTAGTGCAGCGCGGCTACTGGGAGGCTCTGCGCGCAGG 497
QY 661 GCGGATGTTCTGACCTTCTCTGGAGTGTCACTGTGAGTGCCACGAGGCTGCTGGAGCGG 720
Db 498 GCGGAAGTGTGACCTTCTCTAGACTGTCTACTGTGAATGTCTATGAGGGGTGGCTGAGGCT 557
QY 721 CTGCTGAGAGAGTCCATGAAGAGGAGTCCGCGAGTGTGTGCGCGGTGATTGTGATGTCATC 780
Db 558 CTGCTGAGAGAGTCCACGAGAGAGAGTCCGCGAGTGTGTGCGGCTGTATTCGATGTGATT 617
QY 781 GACTTGAACACCTTTGGAATACCTGGGAACTCCCGGAGCGCCCAAGATCGCGGCTTTCGAC 840
Db 618 GACTTGAACACCTTTGAGTACCTGGGCAACTCGGAGAGCGCGAGATTTGGCGGCTTTGAC 677
QY 841 TGGAGGCTGTGTTCAGTGTGCGCACACAGTTCTCTGAGAGGAGAGAGATACGATGCAATCC 900
Db 678 TGGCGGCTGTGTATTTACGTGCGCACGTTTCTCCCGAGCGGAGCGGCGAGTGCATGCGGTG 737
QY 901 CCGTGTGATGTCTACAGTGTCTCCAAACATGCTGTGCGGCTGTTTGTGTGAGTAAGAAA 960
Db 738 CCGATCATGTTATCAGTGTCTCCAACTATGCTGTGGGGCGCTGTTTGTGTGAGTAAGAGA 797
QY 961 TATTTTGAATATCTGGGCTCTTATGATACAGGAATGAAGATTTGGGAGGAGAGAAACCTC 1020
Db -----

```


Db	798	TAITTTGATTATCTGGGGTCTTACGATACAGGAATGGAAGTCTGGGAGGAGAAAACCTT	857
Qy	1021	GAATTTTCCTTTAGGATCTGGCAGTGTGGGGTCTTGAAACACACCCATGTTCCCAT	1080
Db	858	GAGTTCTCTTTAGGATCTGGCAGTGTGGGGTCTTGAAACACACCCATGTTCCCAT	917
Qy	1081	GTTGGCCATGTTTCCCGACGAGCTCCCTACTCCCGACACAGGCTCTGGCCACAGT	1140
Db	918	GTGGGCCACGCTCTTCCCTAAGCAAGCTCCCTATTCCCGTACGAGGCGCTGGCCACAGT	977
Qy	1141	GTTCTGTCAGCTGAAGTATGGATGGATGAATTTAAAGAGCTCTACTACCAATCGCAACCCC	1200
Db	978	GTCCGAGCTGACAGAGTGTGGATGGATGAATTTAAAGAACTCTACTACCAACCCATCCC	1037
Qy	1201	CGTGCCCGCTTGGAACTTTTGGGATGTGACAGAGAGGAGCAGCTCCGGGCAAGCTC	1260
Db	1038	CAGCGCCGCTGGAACTTTTGGGACGTGACAGAGAGGAGAGCTTCGGGCTAAGCTC	1097
Qy	1261	CAGTGTAAAGACTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCGCTGAG	1320
Db	1098	CAGTGTAAAGACTTCAAGTGGTCTTGGAGACTGTGTATCCAGAACTGCATGTGCGCTGAG	1157
Qy	1321	GACAGGCTGGCTCTTCTGGGATGCTCCAGAAACAAAGGACTTAACAGACTACTGCTTTGAC	1380
Db	1158	GACAGGCTGGCTCTTCTGGGATGCTCCAGAAACAAAGGACTTAACAGACTACTGCTTTGAC	1217
Qy	1381	TATAACCTCCCGATGAAACCCAGATTTGTGGGACACCAAGTCACTTGTACCTCTGTCT	1440
Db	1218	TACAATCTCCCAATGAAACCCATGTTGAAAGGCCACCAAGTCTCTGTACCTCTGCCAT	1277
Qy	1441	GGATGGGCGCAGATCAGTTTTTTCGAGTACAGTCCAGAAAGAAATACGCTATAACACC	1500
Db	1278	GGATGGGCGCAGATCAGTTTTTTCGAGTACAGTCCAGAAAGAAATACGCTATAACACC	1337
Qy	1501	CACAGGCTCAGGCTGCTATGCTGTGGAGAGGAGATGATACCCCTTATCATGTCATCTC	1560
Db	1338	CGCAGGCGCCAGGCTGCTATGCTGTGGAGAGGAGATGATACCCCTTATCATGTCATCTC	1397
Qy	1561	TGCGAAGAACTGCCCCAGAGAAATCAGAAGTTTCATCTTGAGGAGGATGATCTTTATTT	1620
Db	1398	TGCGGAGAGACCGTCCAGAGAACCAAGAGTTTCATCTCAGAGGAGTGCACCTTTAGTT	1457
Qy	1621	CAGAACAGTCCAGAAATGTGTCAGGCTGCGAGGAGAGTGCAGTGCATGTTGCTT	1680
Db	1458	CACAAGCAGACGAGGAAATGTGAGGAGCCACAGAGAGTGTAGACAACCGCTTTGCA	1517
Qy	1681	CCACTCTTACGAGCTGCACCACTCGGATCATCAGAAATGTTCTTCAAGAGCGCATG	1740
Db	1518	CCATCTTACGAGCTGTACCACTCAGTAACCAAGGAGTGGTTCTTCAAGGAGCGCATG	1577
Qy	1741	T 1741	
Db	1578	T 1578	
RESULT 6			
AK043977			
LOCUS			
DEFINITION			
AK043977			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			

PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			
TITLE			
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS			


```

Qy 1740 GT 1741
Db 1606 GT 1607

RESULT 7
BE677813
LOCUS
DEFINITION
763d08.x1 Soares NSF F8_9W OT PA P S1 Homo sapiens cDNA clone
IMAGE:3299343 3' similar to TR:008832 008832 POLYPEPTIDE GALNAC
TRANSFERASE-T4. ;contains element MER22 repetitive element ;, mRNA
sequence.
ACCSSON
BE677813 676 bp mRNA linear EST 08-SEP-2000
VERSION
BE677813.1 GI:10038280
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 676)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 477.
Location/Qualifiers
1..676
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3299343"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT7T30-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares Nb2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares Nb2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NbHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 139 a 179 c 221 g 137 t
ORIGIN

Query Match 37.8%; Score 660.8; DB 10; Length 676;
Best Local Similarity 99.0%; Pred. No. 8.4e-138;
Matches 665; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 198 GCGGCGCCGGTCCGCGGAAACGGCGTGGCGCGCGGCGAGCGGTGGCGTGCAGCT 257
Db 1 GCGGCGCCGGTGGCGGCGAAACGGCGTGGCGCGCGGCGAGCGGTGGCGTGCAGCT 60

Qy 258 GCAGGCGAGAGAGTGGCGTGCAGGAGAGAGCGTGCAGTGCACAGATTAACTCTA 317
Db 61 GCAGGCGAGAGTGGCGTGCAGGAGAGAGCGTGCAGTGCACAGATTAACTCTA 120

Qy 318 CCTCAGGACCGGATCTCACTGCACCGCGCGTCCCGAGCGGTGGAAACCGCTGTGCAA 377
Db 121 CCTCAGGACCGGATCTCACTGCACCGCGCGTCCCGAGCGGTGGAAACCGCTGTGCAA 180

378 AGAAGAAATATGATTATGATTAATTTGCCAGGACATCTGTTATCATAGCAATTTTATAA 437
181 AGAAGAAATATGATTATGATTAATTTGCCAGGACATCTGTTATCATAGCAATTTTATAA 240

438 TGAAGCCTGGTCAACTCTCTCCGACAGTTTACAGTGTCTTGGAGACATCCCGGATAT 497
241 TGAAGCCTGGTCAACTCTCTCCGACAGTTTACAGTGTCTTGGAGACATCCCGGATAT 300

498 CTGCTAGAGAAGTGTATCTCTAGATGACTACAGTGTAGAGAGACACCTGAAGGAGCG 557
301 CTGCTAGAGAAGTGTATCTCTAGATGACTACAGTGTAGAGAGACACCTGAAGGAGCG 360

558 CTTGGCCAATAGACTTTTCGGGACTGCCCAAGGTGCGCTGATCCGCGCCAAACAGAGA 617
361 CTTGGCCAATAGACTTTTCGGGACTGCCCAAGGTGCGCTGATCCGCGCCAAACAGAGA 420

618 GGGCCTGGTGGAGCCCGGCTGCTGGGGGGCTGCTGGGGCGAGGGCGATGTTCTGACCTT 677
421 GGGCCTGGTGGAGCCCGGCTGCTGGGGGGCTGCTGGGGCGAGGGCGATGTTCTGACCTT 480

678 CTGGAAGTGTCACTGTGAGTGCACCAAGGCTGGAGCCGCTGCTGCAGAGGATCCA 737
481 CTGGAAGTGTCACTGTGAGTGCACCAAGGCTGGAGCCGCTGCTGCAGAGGATCCA 540

738 TGAAGAGGAGTTCGGCAGTGTGTGCCCGGTGATTGATGATCGACTGGAAACCTTCGA 797
541 TGAAGAGGAGTTCGGCAGTGTGTGCCCGGTGATTGATGATCGACTGGAAACCTTCGA 600

798 ATACCTGGGGAATCCGGGAGAGCCAGATCGCGGTTTCGACTGGAGGCTGTGTTCAC 857
601 ATACCTGTGGAACTCCGGGAGAGCCAGATCAGCGGTTTCGACTGGAGGCTGTGTTCAC 660

858 GTGGCACACAGT 869
661 GTAGCACACAGT 672

RESULT 8
CB165548/c
LOCUS
DEFINITION
BEQ603160969.R1 CSEQFXN34 liver Bos taurus cDNA, mRNA sequence.
CB165548
ACCESSION
CB165548.1 GI:28151673
VERSION
EST.
KEYWORDS
SOURCE
Bos taurus (cow)
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 692)
Adelson, D.L. and Gill, C.A.
AUTHORS
Bovine ESTs (Adelson and Gill)
TITLE
Unpublished
JOURNAL
COMMENT
Contact: David L. Adelson
Animal Breeding and Genetics
Texas A&M University
Animal Science Dept., TAMU-2471, College Station, TX 77843-2471,
USA
Tel: 9798452616
Fax: 9798456970
Email: david.adelson@tamu.edu.
FEATURES
source
1..692
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="liver"
/clone_lib="CSEQFXN34 liver"
/notes="Organ: liver; Vector: pBluescript SK+; Site_1: NotI
; Site_2: EcoRI; sequence 5' of the insert
(5'-NNN...NNNinsert)
GCCAATTGGAGCTCCACCAGCGGTGGCGGCGGCTCGAG. Sequence 3' of

```

the inserts (AAGAAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAG.
normalized Rd 2 library, sequenced 3' with M13R primer. "

BASE COUNT	149 a	214 c	174 g	155 t	ORIGIN
Query Match	32.4%	Score 566.4	DB 14	Length 692	
Best Local Similarity	89.6%	Pred. No. 1.4e-116			
Matches 620; Conservative	0	Mismatches 71	Indels 1	Gaps 1	
Qy	330	CATCTCACTGCACCGCCGCTGCCGAGCGGTGGAACCGCTGTGCAAAGAGAAGAATA	389		
Db	692	CATCTCGCTGCACCGCCGCTGCCGAGCGGTGGAACCGCTGTGTGCAAGAGAAGAATA	633		
Qy	390	TGATTATGATTAATTTGCCAGGACATCTGTTATCATAGCATTTTATATGAAGCCTGGTC	449		
Db	632	TAAGTAGCATGAATTTGCCACGACATCTGTCTCATAGCGTTTATATGAAGCCTGGTC	573		
Qy	450	AACTCTCTCTCGGACAGTTTACAGTGTCTTGGACATCCCGGATATCCCTGCTAGAGA	509		
Db	572	AACTCTCTCTCGGACAGTTTACAGTGTCTCTCGAGACATCCCGGACATCCCTGCTAGAGA	513		
Qy	510	AGTGATCTTTGTAGATGACTACAGTGTAGAGACACCTGAAGAGCGCTTTGGCCAAATGA	569		
Db	512	AGTTATCTTTGTAGATGACTACAGTGTAGAGACACCTGAAGAGCGCTTTGGCCAAATGA	453		
Qy	570	GCTTTTGGGACCTGCCAAGTGTGCGCTGTATCCGCGCCAAAGAGAGAGGGCGCTGTGGCG	629		
Db	452	GCTGGCGGGTGTGCCAAGTGTCCGCTGTATCCGAGCCAAAGAGAGGGCGCTGTGGCG	393		
Qy	630	AGCCCGGCTGTGGGGGCGTCTGGCGGAGGGCGATGTTCTGACCTTCCTGGACTGTCA	689		
Db	392	AGCACGGTGTCTGGGGGCGTCTGGCGGAGGGCGAGTGTGACCTTTCTGGACTGTCA	333		
Qy	690	CTGTGAGTGCCAACGAAGGTGGCTGGAGCCGCTGCTGCAGAGGATCCATGAAGAGGAGTC	749		
Db	332	CTGCGAGTGCCATGAGGGGTGGCTGTGAGCCACTGCTGCAGAGGATCCATGAAGAGGATC	273		
Qy	750	GGCAGTGGTGTGCCCGGTGATTTGATGTGATCGACTGGAAACCTTCGAATACCTGGGAA	809		
Db	272	AGCGGTGGTGTGCCCGGTGATTTGATGTGATCGACTGGAAACCTTCGAATACCTGGGAA	213		
Qy	810	CTCCGGGAGGCCACAGATCGG - CGGTTTCGACTGGAGCGCTGGTGTTCAGTGGCACACAG	868		
Db	212	TGCTGGGGAGGCCACAGATCGGACGGTTTGTGACTGGCGGCTGGTGTTCAGTGGCACACGG	153		
Qy	869	TTCTGTGAGAGGGAGAGGATACGGATGCAATCCCGCTCGATGTTCATCAGGTCTCCAAACAA	928		
Db	152	TTCCGAAAGAGAGAGGATCCGATTCGGGTCCCCCATCGATGTTCATCAGATCTCCACGA	93		
Qy	929	TGGCTGTGGGCTGTTTCTGTGAGTAAGAAATATTTTGAATATCTGGGGTCTTATGATA	988		
Db	92	TGGCTGTGGGCTGTTTCTGTGAGTAAGAAATCTTTGAATACCTGGGGTCTTATGATA	33		
Qy	989	CAGGAATGGAAATTTGGGGAGGAGAAACCTTC	1020		
Db	32	TAGGAATGGAAATTTGGGGTGGAGAAACCTTC	1		

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 915 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 389.
Location/Qualifiers
1. .561
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2365083"
/lab_host="DH10B"
/clone_lib="Soares NSF_F8_9W_OT_PA_P_S1"
/notes="Organ: pooled; Vector: p7Y73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and cloneIDs: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares Nb2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares Nb2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bencio
Soares and M. Patricia Ronaldo "

BASE COUNT	ORIGIN	121 a	144 c	177 g	119 t	Source and M. of Origin
Query Match		31.4%	Score 548.2;	DB 9;	Length 561;	
Best Local Similarity		98.6%;	Pred. No. 1.5e-112;			
Matches 553; Conservative		0;	Mismatches 8;	Indels 0;	Gaps 0;	
Qy	241	GC	GGTGGGCTGCAGCTGCAGGGCGAGGAGCTCGGGCTGCAGAGGAGACGGTGGCGCTG	300		
Db	1	GC	GGTGGGCTGCAGCTGCAGGGCGAGGAGCTCGGGCTGCAGAGGAGACGGTGGCGCTG	60		
Qy	301	CAC	CAGATTAACTACCTCTACGCGACCGCATCTCACTGCACGCCCGCCCTGCCGAGCGC	360		
Db	61	CAC	CAGATTAACTACCTCTACGCGACCGCATCTCACTGCACGCCCGCCCTGCCGCTGCGC	120		
Qy	361	TG	GAACCGGCTGTGCAAGAGAGAAATATGATATGATTAATTTGCCAGGACATCTGTT	420		
Db	121	TG	GAACCGGCTGTGCAAGAGAGAAATATGATATGATTAATTTGCCAGGACATCTGTT	180		
Qy	421	AT	CATAGCATTTTATTAATGAAGCCTGGTCAACTCTCCTTCGGAAGTTTACAGTGTCTT	480		
Db	181	AT	CATAGCATTTTATTAATGAAGCCTGGTCAACTCTCCTTCGGAAGTTTACAGTGTCTT	240		
Qy	481	GAG	ACATCCCGGATATCCTGTAGAGAAAGTAGATCCTTGTAGATGACTACAGTGATAGA	540		
Db	241	GAG	ACATCCCGGATATCCTGTAGAGAAAGTAGATCCTTGTAGATGACTACAGTGATAGA	300		
Qy	541	GAG	CACTGAAGAGCGCTTTGGCCAAATGAGCTTTGGGACTGCCCAAGGTGGCGCTGATC	600		
Db	301	GAG	CACTGAAGAGCGCTTTGGCCAAATGAGCTTTGGGACTGCCCAAGGTGGCGCTGATC	360		
Qy	601	CG	CCCAAGAGAGAGGGCGCTTGGTCGAGCCCGGCTGTCTGGGGGGCTCTGCGCGCAGG	660		
Db	361	CG	CCCAAGAGAGAGGGCGCTTGGTCGAGCCCGGCTGTCTGGGGGGCTCTGCGCGCAGG	420		
Qy	661	GG	CGATGTTCTGACCTTCTGGAATGATGCACTGTGAGTGCACGAAGGGTGGCTGGAGCGG	720		
Db	421	GG	CGATGTTCTGACCTTCTGGAATGATGCACTGTGAGTGCACGAAGGGTGGCTGGAGCGG	480		

Qy	721	CTGCTGCAGAGGATCCATGAAGAGGATCGGCACTGTGTGCGCGGTGATTGATGATC	780
Db	481	CTGCTGCAGAGGATCCATGAAGAGGATCGGCACTGTGTGCGCGGTGATTGATGATC	540
Qy	781	GACTGGGAACCTTCGGAATAC	801
Db	541	GACTGGGAACCTTCGGAATAC	561
RESULT 10			
LOCUS	AK033494	2607 bp	linear
DEFINITION	Mus musculus adult male colon cDNA, RIKEN full-length enriched library, clone:9030401N20 product:UDP-N-acetyl-alpha-D-galactosamin 6-polyphosphate N-acetyl-galactosaminyltransferase 4, full insert sequence.		
ACCESSION	AK033494		
VERSION	AK033494.1	GI:26329156	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitagawa, T., Tashiro, H., Itoh, M., Sugi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishigaki, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, F., Kaishiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kikawa, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		
REFERENCE	4		
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischer, C., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nishida, I., Pebole, G., Quackenbush, J., Schiraldi, L. M., Stauber, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409 (6821), 685-690 (2001)		
MEDLINE	21085660		

PUBMED
REFERENCE
AUTHORSTITLE
JOURNAL
REFERENCE
AUTHORSTITLE
JOURNAL

COMMENT

FEATURES
source

CDS

polya_signal
polyA_site
BASE COUNT 677 a 628 c 660 g 642 t

11217851

5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 2607)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawaji, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

Location/Qualifiers

1. 2607

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM DB:9030401N20"

/db_xref="taxon:10090"

/clone="9030401N20"

/sex="male"

/tissue_type="colon"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

246. 1982

/notes="unnamed protein product;
UDP-N-acetyl-alpha-D-galactosamine: polypeptide
N-acetyl-galactosaminyltransferase 4 (MGI:894692,
GB|NM_015737, evidence: BLASTN, 100%, match=1737)
putative"

11217851

5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 2607)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawaji, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

Location/Qualifiers

1. 2607

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM DB:9030401N20"

/db_xref="taxon:10090"

/clone="9030401N20"

/sex="male"

/tissue_type="colon"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

246. 1982

/notes="unnamed protein product;
UDP-N-acetyl-alpha-D-galactosamine: polypeptide
N-acetyl-galactosaminyltransferase 4 (MGI:894692,
GB|NM_015737, evidence: BLASTN, 100%, match=1737)
putative"

ORIGIN

Query March	29.8%	Score 520.8;	DB 11;	Length 2607;
Best Local Similarity	63.1%;	Pred. No. 4.1e-106;		
Matches 820;	Conservative 0;	Mismatches 477;	Indels 3;	Gaps 1;
Qy	201	GCGCCCGGTGCGCGGAACGCGCTGGCGCGCGCGGGCGAGCGGTGCGGCTGCAGCTGCA	260	
Db	443			
Qy	261	GGCGAGAGCTGCGGTGCGAGGAGAGCGTGCCTGCGCTGCAACAGATTAACTATCACT	320	
Db	503	CGAGGGGAACTGAAGCAGCAGAAGAAACTTATTAGAGATATGCCATCAACATTTACCT	562	
Qy	321	CAGCGACCGCATCTCACTGCACCGCGCTGCCCGAGCGCTGGNAACCGCTGTGCAAGA	380	
Db	563	CAGTGAAGGATCTCCCTGCAACCGCCACATAGAGGATAAAGAAATGTATGAGTGAAGC	622	
Qy	381	GAAGAAATATGATTATGATMAATTGGCCAGGACATCTGTTATCATAGCATTTTATAATGA	440	
Db	623	CAAGAAGTCCACTACAGGTCGCTCCCAACCACTCCGTTATCATCGCTTCTATAACGA	682	
Qy	441	AGCTGTGTAACCTCTCTTTCGGAAGTTTACAGTGTCTTTGAGACATCCCGGATATCCT	500	
Db	683	AGCTGTGTCCACTCTGCTTCGGACCAATTCACAGTGTGTTTGAAGAACTTCTCCAGCTGTCT	742	
Qy	501	GCTAGAGAAGTGTATCTTGTGTAGATGACTACAGTGTAGAGAGCACCTGAAGGAGCGCTT	560	
Db	743	TTTAAAGGAGTCACTCTTGTGTGATGACTTGAGTGTATAGAAATTTATTTGAAGCGCGAACT	802	
Qy	561	GGCCAAATGAGCTTTTCGGACATGCCCAAGGTGCGCTGATCGCGCCCAACAGAGAGAGGG	620	
Db	803	TGAAACTTACATCAGCAACCTGGAGAGAGTTCGCTTGATTAGAACCAATAAGAGAGAGGG	862	
Qy	621	CCTGGTGGAGCCGGCTGCTGGGGCGTCTGCGCGAGGGCGCATGTTCTGACCTTCCT	680	
Db	863	GCTGGTCCGGCCCGCTCTAAATGGGGCCACTTTTGCCACCGGGGATGTCTCATCTTCCT	922	
Qy	681	GGACTGTCACTGTGAGTGCCACGAAGGTGGCTGGAGCCGCTGCTGCAGAGGATCCATGA	740	
Db	923	GGATTGTCACTGTGAGTGTAACACTGGTTGGTTGGAACCACTTTTGGAGAGGATCAGTAG	982	
Qy	741	AGAGAGTCGCGAGTGTGTGTCGCCGGTGATTTGATGTGATCGACTGGAAACACCTTCGAATA	800	
Db	983	AGACGAAACCGCAATCGTATGCTCTGTCTATCGACACTATTGACTTGGAACTTTTCSAAT	1042	
Qy	801	CCTGGGGAACCTCGGGAGCCCCAGATCGGCGGTTTGCAGCTGGAGGCTGGTTTCACTG	860	
Db	1043	CTACATGCAAGCCGGGAGCCCATGATTTGGTGGGTTTGACTGGGGTTTGACCTTCCAGTG	1102	
Qy	861	GCACACAGTTCTTGAGAGGAGAGGATACCGATGCAATCCCCCGTCGATGTTCATCAGGTC	920	
Db	1103	GCAATTCGTCCCCAAACATGAAGAGGACAGGCGAAAGTCAAGAAATGACCCCATCAGGTC	1162	
Qy	921	TCCAAACATGCTGGTGGCTGTTTGTGTGAGTAAGAAATATTTTGAATATCTGGGGTC	980	
Db	1163	CCCCACATGCGAGAGGACTGTTTGTGTGTGAGCAAGAGTATTTTTCAGTACCTTGGGCA	1222	
Qy	981	TTATGATACAGAAATGGAAGTTTGGGAGGAGAAAACTCGAAATTTTCTTTAGGATCTG	1040	
Db	1223	CTATGACACTGGCATGGAAGTCTGGGAGGCGAAAACTTGAGTTGCTTTCAGGGTTG	1282	
Qy	1041	GCAGTGTGGTGGGTTCTGGAAAAACACCAATGTTTCCCATGTTGGCCATGTTTTTCCCAA	1100	
Db	1283	GCAGTGTGGCGGCAAGCTGGAGATCCACCCATGCTCGCACGTGGGCCCACTGTTTCCCAA	1342	
Qy	1101	GCAAGCTCCCTACTCCGCGCAACAAAGGCTCTCGCCAAACAGTGTCTGTGCGAGCTGAAGTATG	1160	
Db	1343	CGCGCGGCCCTATGCTGCCCCCACTTCTACAGAAATACCGCGGGCGGCGGAGGTG	1402	
Qy	1161	GATGGATGAATTTTAAAGAGCTCTACTACCATTCGCAACCCCGTGCCCGCTTGGAACTTTT	1220	
Db	1403	GATGGATGAATACAAAGAGCATTTCTATAACCGAAACCCCGCCAGCTCGGAAAGAGGCTTA	1462	

Qy	1221	TGGGGATGTGACAGAGAGGAAGCAGCTCCGGGACACAGCTCCAGCTGTAAAGACTTCAAGTG	1287
Db	1463	TGGCGATCTTTCTGAAAGGAAGTTACTTAAGGGGAACGGCTGAAAATGTAAGAGCTTCGACTG	1522
Qy	1281	GTTCCTGGGACACTGTGTATCCAGAACTGCGATGTCCTTGAGGACAGGGCTTCGGCTTCTTCGG	1340
Db	1523	GTATCTGAAAANAATGTGTCTTCTTAACCTTACACGTCCCGAGGATAGACCTGGCTGGCACGG	1582
Qy	1341	GATGCTCCGAGAACAAAGGAGCTAAACAGACTACTGCTTTGACTATATAACCTCCCGATGAAAA	1400
Db	1583	AGCATCCGACGATGGGCATCTCTCCGAGTGCTTAGATTATAACGCTCCCGACAAACA	1642
Qy	1401	CCGAGTTGGGACACACAGGTCAATCTGTACCTCTGTGATGGGATGGCCAGCAATCAGTT	1460
Db	1643	CCCTACGGGGGCTAAATC---TTTCGTGTGTTGGATGCCACGGTCAAGGAGGCAATCAGTT	1699
Qy	1461	TTTCGAGTACACGTCCTCCAGAAAGAAATACGCTATAACACC	1500
Db	1700	CTTTGGAATATACTTCAACAAGAANAATCAGATCAATCC	1739
RESULT 11			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
Qy			
Db			


```
QY 258 GCAGGCGAGAGCTGGCGCTGCAGGAGGAGCGTGGCGCTGCACGAGTAAATCACTCTA 317
Db 61 GCAGGCGAGAGCTGGCGCTGCAGGAGGAGCGTGGCGCTGCACGAGTAAATCACTCTA 120
QY 318 CCTCAGGACCGCATCTCACTGCACCGCGCTGCCGAGCGCTGGAACCGCTGTGCAA 377
Db 121 CCTCAGGACCGCATCTCACTGCACCGCGCTGCCGAGCGCTGGAACCGCTGTGCAA 180
QY 378 AGAGAAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATFAGCATTTTATAA 437
Db 181 AGAGAAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATFAGCATTTTATAA 240
QY 438 TGAAGCTGTGTCACCTCTCTTCCGACAGTTTACAGTGTCTTGAAGACATCCCCGGATAT 497
Db 241 TGAAGCTGTGTCACCTCTCTTCCGACAGTTTACAGTGTCTTGAAGACATCCCCGGATAT 300
QY 498 CCTCTAGAGAGAGTGTCTCTTGTAGATGACTACAGTGATAGAGAGCACCTGGAAGGAGCG 557
Db 301 CCTCTAGAGAGAGTGTCTCTTGTAGATGACTACAGTGATAGAGAGCACCTGGAAGGAGCG 360
QY 558 CTTCGCCAATAGAGCTTTGGGAGCTGCCAAGGTGGCTGTATCCGCGCAACAAAGAGAGA 617
Db 361 CTTCGCCAATAGAGCTTTGGGAGCTGCCAAGGTGGCTGTATCCGCGCAACAAAGAGAGA 420
QY 618 GGG-CCTGTGTGAGAGCGCGCTGTGGGGCGTCTGCGGCGAGGGCGATGTTCTGACCT 676
Db 421 GGG-CCTGTGTGAGAGCGCGTGTGGGGCGTCTGCGGCGAGGGCGATGTTCTGACCT 480
QY 677 TCCTGGAGTGTCACTGTGAGTGCCACGAGGG 708
Db 481 T-CTGGAGTGTCACTGTGAGTGCCAGAAAGGG 511
```

```
RESULT 15
AA429394
LOCUS
DEFINITION
w32h12.81 Soares ovary tumor N8HOT Homo sapiens cdna clone
IMAGE:771047 3', similar to TR:E209711 E209711
UDP-GALNAC:POLYPEPTIDE N-ACETYLGLACTOSAMINYLTRANSFERAS. ;, mRNA
sequence.
AA429394
AA429394.1 GI:2112349
EST.
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 496)
Hallier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,
T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
Unpublished
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyT not found
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 310.
Location/Qualifiers
1..496
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:5980921"
/db_xref="taxon:9606"
```

FEATURES

source

```
/clone="IMAGE:771047"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares ovary tumor N8HOT"
/notes="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cdna was primed with a Not I - oligo (dT) primer [5'-
TGTTCACCAATCTGAAGTGGAGGCGCGGCTTTTTTTTTTTTTTTT 3'],
double-stranded cdna was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo."
```

BASE COUNT 107 a 132 c 157 g 100 t

```
Query Match 25.8%; Score 450.8; DB 9; Length 496;
Best Local Similarity 96.8%; Pred. No. 1.1e-90;
Matches 482; Conservative 0; Mismatches 12; Indels 4; Gaps 2;

QY 203 CGCGCGTGGCGGGAACGCGCTGGCGCGCGGCGGCGGCTGCGGCTGCAGCTGCAGG 262
Db 2 CGCGCGTGGCGGGAACGCGCTGGCGGCGGCGGCGGCTGCAGCTGCAGG 58
QY 263 GCGAGGAGCTGCGGCTGCAGGAG-GAGAGCGTGGCGTGCACGAGTAAATCACTC 321
Db 59 GCGAGGAGCTGCGGCTGCAGGAGGAGCGTGCAGCTGCACGAGTAAATCACTC 118
QY 322 AGCGACCGCATCTCACTGCACCGCGCTGCCGAGCGCTGGAACCGCTGTGCAAGAG 381
Db 119 AGCGACCGCATCTCACTGCACCGCGCTGCCGAGCGCTGGAACCGCTGTGCAAGAG 178
QY 382 AAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTATAATGAA 441
Db 179 AAGAAATATGATTATGATAATTTGCCAGGACATCTGTTATCATAGCATTTATAATGAA 238
QY 442 GCCTGTGTCAACTCTCCTTCGAGCAGTTTACAGTGTCTTGGAGACATCCCGGATATCCTG 501
Db 239 GCCTGTGTCAACTCTCCTTCGAGCAGTTTACAGTGTCTTGGAGACATCCCGGATATCCTG 298
QY 502 CTAGAGAGAGTGTATCTTGTAGATGACTACAGTGATAGAGAGCACCTGAAGGAGCGCTTG 561
Db 299 CTAGAGAGAGTGTATCTTGTAGATGACTACAGTGATAGAGAGCACCTGAAGGAGCGCTTG 358
QY 562 GCCAATGAGCTTTCCGAGCTGCCAAGGTGGCGCTGATCCGCGCAACAAAGAGAGAGGCG 621
Db 359 GCCAATGAGCTTTCCGAGCTGCCAAGGTGGCGCTGATCCGCGCAACAAAGAGAGAGGCG 418
QY 622 CTGCTGCGAGCCCGGCTGTGGGGGGCTCTGCGGCGAGGGCGGATGTTCTGACCTTCTG 681
Db 419 CTGCTGCGAGCCCGGCTGTGGGGGGCTCTGCGGCGAGGGCGGATGTTCTGACCTTCTG 478
QY 682 GACTGTCACTGTGAGTGC 699
Db 479 GACTGTCACTGTGAGTGC 496
```

Search completed: December 7, 2003, 21:17:09
Job time : 3640.14 secs

THIS PAGE BLANK (USPTO)